Cosima Schneider[1], Markus Janczyk[2] & Nadine Bade

2022

Verifying Pragmatic Inferences in Context: An Experimental Comparison of Presuppositions of *again* and the Definite Determiner with Scalar Implicatures

[1]ORCID: 0000-0002-4314-0885; [2]ORCID: 0000-0002-9958-3220

in

Robin Hörnig, Sophie von Wietersheim, Andreas Konietzko & Sam Featherston (eds.),

Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data

pp. 79–95

Tübingen: University of Tübingen

https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301

In R. Hörnig, S. von Wietersheim, A. Konietzko & S. Featherston (eds.) (2022), *Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data*. Tübingen: University of Tübingen. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301

Verifying Pragmatic Inferences in Context: An Experimental Comparison of Presuppositions of *again* and the Definite Determiner with Scalar Implicatures

Cosima Schneider, Markus Janczyk & Nadine Bade
University of Tübingen, University of Bremen, University of Potsdam
cosima.schneider@uni-tuebingen.de

1 Introduction

Presuppositions are ubiquitous in everyday life and a vital topic in the semantic and pragmatic literature (e.g., Beaver & Geurts, 2012, for an overview). To gain a deeper understanding of the phenomenon, psycho-linguistic methods have been adopted to study how presuppositions are processed. Previous evidence suggests that presuppositions are processed immediately, the process starting already on the presupposition trigger (e.g., Schwarz, 2007; Tiemann et al., 2011; Kirsten et al., 2014; Schneider et al., 2019). Further studies suggest that different types of presupposition triggers differ in their processing (Domaneschi et al., 2014; Tiemann et al., 2015; Domaneschi & Di Paola, 2018). The present paper investigates a fundamental issue of presupposition processing, that is, the validation of presuppositions in a given context. To that end, we compare test sentences that trigger presuppositions in contexts that either falsify or verify them. We focus on two particular triggers, namely the definite determiner *the* (German *der/die/das*) and *again* (German *wieder*).

Overall, we pursue four goals. First, we aim to replicate results indicating an early processing of presuppositions, starting as soon as the presuppositional content is fully known (e.g., Tiemann et al., 2011), with a larger sample size. Second, we aim to investigate classifications of triggers offered in the theoretical literature by comparing two presupposition triggers assumed to belong to different classes. Third, we investigate how scalar implicatures (SI) are processed. Fourth, we compare the time-course of processing presuppositions and scalar implicatures in verifying/falsifying contexts to further explore the current theoretical distinction of different types of pragmatic content more generally.

The paper is structured as follows. In Section 2, we introduce the theoretical and experimental background for presuppositions and scalar implicatures we assume. Section 3 introduces the idea behind and the procedure of our experiment. Section 4 describes the method in detail. Results are presented in Section 5 and discussed in Section 6. Section 7 concludes the paper.

2 Theoretical Approaches to Presuppositions and Implicatures

2.1 Presuppositions

Theoretical approaches to presuppositions. According to a traditional **semantic approach**, presuppositions are encoded in the lexical entry of a presupposition trigger, that is, in a lexical item. Presuppositions are regarded as "definedness conditions": a sentence *s* presupposes *p* just in case *p* must be true in order for *s* to have a truth value at all (Strawson, 1952). If the presupposition is not met, the sentence is undefined rather than true or false (Heim & Kratzer, 1998). Formally, this is captured by making sentences with presuppositions denote partial functions.

An example for a presupposition trigger is *again*, eliciting the presupposition that the event it combines with happened before the actual utterance. Accordingly, a sentence like (1) is considered true if Anna scored and scored before, false if she did not score but scored before, and undefined otherwise.

(1) Anna scored, again.

More formally, (1) is a function from worlds to truth values (propositions) only defined for those worlds where Anna scored before. Semantic undefinedness is mapped onto pragmatic inappropriateness through "Stalnaker's Bridge" (von Fintel, 2004). As a result, uttering a sentence with a presupposition in a context that does not satisfy it will yield pragmatic oddness.

At least some presuppositions have been argued to have a pragmatic source though (Stalnaker, 1973; Simons, 2001; Abusch, 2002, 2010; Schlenker, 2012; Abrusán, 2011), that is, they are not conventionally (i.e., lexically) encoded but are derived contextually. We refer to this idea as the **pragmatic approach**. Based on different triggering mechanisms, Abusch (2002) introduced the distinction of *hard* and *soft* presupposition triggers (see also Abrusán, 2011; Romoli, 2011). The two types behave differently in contexts where the speaker is ignorant with regard to the truth of the presupposition. For example, *soft* triggers can be uttered in explicit ignorance contexts[1] without leading to infelicity, but *hard* triggers cannot be. For example, (2) with the *soft* trigger *stop* is fine, whereas (3) with the *hard* trigger *too* is odd.

(2) I don't know if John ever smoked, but if he recently stopped that would explain his nervous behavior.

(3) #I don't know if anyone came, but if Peter came, too, that would explain why Mary is not happy.

Other examples for *soft* triggers are *continue* (see Simons, 2001), *know*, and achievement verbs such as *win*. *Hard* triggers are, for example, *too*, *also*, *even*, *again*, the negative polarity item *either*, and *it-clefts* (see also Abusch, 2010). *Soft* triggers share certain properties like suspendability and non-detachability with conversational implicatures. Based on that observation, Romoli (2011) proposed that presuppositions of *soft* triggers *are* SIs, which come out as entailments in certain environments. It is still under debate which trigger belongs to which class, and – relatedly – which triggers may have a conversational source.

From both the pragmatic and semantic approach one can derive the hypothesis that the listener is confronted with increased cognitive workload in cases where the presupposition is false or unfulfilled. However, the approaches diverge in their predictions about **when** this increase becomes apparent: for one, and according to the semantic approach, the presupposition is encoded in the trigger. Difficulties are expected to arise upon encountering the trigger and should persist until the content of the presupposition is known. In contrast, and according to the pragmatic approach, the sentence is first interpreted compositionally, and only then is the context checked for whether it is appropriate given its presupposition. The distinction need not be this clear cut though, and it is possible to assume that pragmatic information is taken into consideration at every given point during computation, for example, with the assumption of local contexts (see Schlenker, 2011).

Experimental evidence. In recent years, an increasing amount of experimental studies on presupposition processing has been carried out (see Schwarz, 2016 for a review). Results from some studies support the semantic approach. For example, Tiemann et al. (2011) compared reading times of test sentences including a presupposition trigger with grammatical sentences without a presupposition and with ungrammatical sentences. Directly on the trigger, reading

[1] Simons (2001) introduced the term *explicit ignorance contexts* to describe situations in which it is apparent to the addressee that the speaker is ignorant with respect to the proposition that would normally be presupposed.

times were longest in the trigger condition, followed by the neutral condition, while the unacceptable condition evoked the shortest reading times (see also Schneider et al., 2020b). This suggests that the trigger requires more attention, possibly because it alerts the reader to look back at the preceding context. These early effects support the idea of an immediate processing of presuppositions starting on the trigger itself.

Schwarz (2007) also reported results from self-paced reading studies for the trigger *also* in both German and English when the presupposition was either met or not supported within the presented sentence. The data revealed longer reading times on the region containing *also* in contexts that did not satisfy the presupposition compared to contexts that satisfied it. This is evidence for an immediate processing of presupposition, which is assumed to be more difficult in cases when the presupposition is not supported by the context.

Similar results are provided by van Berkum et al. (1999) who investigated referentially ambiguous noun phrases. They observed early effects when the uniqueness presupposition of the definite determiner was not met (see also van Berkum et al., 2003). The data show that people realize the presupposition failure of a definite noun phrase immediately on the noun phrase itself. The data also show that this process is more difficult than the verification of the presupposition. Additionally, Tiemann (2014) and Tiemann et al. (2015) discussed reading time studies that reveal processing difficulties on the trigger when the presupposition is not met in the context.

Further support for an immediate processing of presupposition triggers comes from EEG studies by Kirsten et al. (2014). These authors investigated the processing of definite and indefinite determiners in two types of context sentences that either introduced a single object (e.g., one polar bear) or multiple objects (e.g., some polar bears). The data revealed that participants recognized the mismatching condition already when reading the determiner. Burkhardt's (2006) ERP study further supports the idea of early processing of presuppositions by revealing an N400 effect on the trigger position when the existence presupposition of the definite determiner was not fulfilled.

Further evidence for an immediate processing of presuppositions comes from a visual-world eye-tracking study using a picture selection task (Bade & Schwarz, 2019) where participants looked at the respective target picture very early upon hearing the noun. This suggests that the information about uniqueness or anti-uniqueness encoded in determiners was used rapidly for interpretation of the test sentence, already while hearing the determiner itself (see also Schneider et al., 2020a for a similar conclusion from a mouse-tracking study). Additional support for an early processing of presupposition comes from a mouse-tracking study by Schneider et al. (2019) where participants were instructed to evaluate test sentences in a certain context and judge them as appropriate or inappropriate via moving the mouse cursor into the corresponding response boxes in the upper right/left corner of the screen.

While some of the aforementioned studies focused on one or two presupposition triggers, other studies also suggest that different types of presupposition triggers may differ in processing (Domaneschi et al., 2014; Domaneschi & Di Paola, 2018). For example, Domaneschi et al. (2014) suggest that speakers have a different attitude towards the processing of a presupposition when the context does not support it. A mandatory processing of presupposed information is induced by factive verbs, whereas iteratives or focus sensitive particles lead to optional processing of the information. Further, in case of presupposition accommodation, Domaneschi & Di Paola (2018) observed longer processing times for iteratives than for focus sensitive particles and change of state verbs. Along these lines, Tiemann (2014) discussed *the* and *again* as falling into two different classes, especially regarding accommodation. Class 1 triggers are non-informative triggers (Tonhauser, 2015) whose presupposition and assertion are not dependent on each other. Tiemann argued that, because of this independence, the presupposition can be ignored (need not be accommodated) for evaluating the presupposition. In contrast to that, presuppositions of Class 2 triggers are crucial prerequisites for judging the truth of the assertion.

For example, the existence presupposition of the definite must be verified before the existential quantification that is part of the assertive meaning contribution can be evaluated. Tiemann presented several empirical arguments for this division. There is an overlap with the *soft*/*hard* distinction as all Class 1 triggers are *hard* triggers. However, there are also *hard* triggers that are in Class 2. Both categorizations must be considered broad, as more fine-grained differences have been observed, and, as a result, more complex taxonomies have been proposed (Simons et al., 2010).

2.2 Scalar Implicatures

Another prominent example of pragmatic inferences are SIs, which are theoretically distinguished from presuppositions based on their properties. An example is given in (4).

(4) Some elephants are mammals.
SI: Not all elephants are mammals.

Theoretical approaches to scalar implicatures. Under a **Gricean** (or pragmatic) **view** (Grice, 1989), implicatures arise as a result of pragmatic reasoning based on the *Cooperative Principle* and the four Gricean maxims of conversation. The scalar implicature in (4) is the result of reasoning with quantity: as the sentence with *all* is strictly more informative, the speaker deduces that the hearer would have used it if s/he believed it to be true. As s/he did not, the hearer can safely assume that the speaker does not believe the *all* alternative to be true. According to Grice, the process is highly context-dependent, and the resulting inferences are weak and cancelable.[2] Under a **lexical approach** (Levinson, 2000; Chierchia, 2004), a lexical ambiguity is essentially argued for. The implicature is added to the lexical meaning of the scalar term in such a way that *some* has a meaning *some and not all*, which is claimed to be more efficient than assuming reasoning over alternatives in several steps. Levinson (2000) proposed that the scalar implicature is marked as defeasible in the lexicon, whereas Chierchia (2004) slightly deviated from Levinson and introduced a scalar term with a *weak* (implicature-free) meaning and with a *strong* (implicature-laden) meaning.

To explain how listeners arrive at an interpretation of the pragmatic *some* (*some, but not all*), Huang & Snedeker (2009) introduced a processing model based on the Gricean and lexical approach. This "literal-first view" assumes that the lower-bound semantic interpretation (*some and possibly all*) is computed rapidly as part of the basic sentence meaning. All inferences including a scalar implicature and the upper-bound meaning require extra time and resources, and processing takes place in two steps: first, the semantic meaning (*some and possibly all*) is constructed, and second, the pragmatic meaning (*some, but not all*) is computed. Two steps are necessary because the pragmatic interpretation cannot exist without the semantic one. Accordingly, the pragmatic (or upper-bound) meaning of *some* requires more processing resources and thus takes longer than the processing of the semantic (or lower bound) meaning of *some*.

Finally, there is – besides the Gricean and the lexical approach – the **grammatical approach** to scalar implicatures that suggests scalar implicatures arise as entailments of exhaustified sentences (see, e.g., Chierchia, 2006; Fox, 2007; Magri, 2010). The exhaustivity operator EXH that derives the exhaustification is similar in meaning to overt *only* (modulo presuppositions). It applies to a proposition *p* and its alternatives and affirms the proposition while negating the subset of relevant alternatives.

In terms of a processing model, this idea is known as the "Default view" (Levinson, 2000). It predicts that the processing of scalar implicatures is effortless and immediate because of their status as default inferences. Only cancellation of the implicature incurs processing resources

[2] Under so called Neo-Gricean approaches (Sauerland, 2004), the basic idea of scalar implicatures as the result of pragmatic reasoning is kept, but a strengthening mechanism is introduced, as well as the notion of local implicatures. We won't discuss this approach in much detail here, as it is not relevant for the predictions we tested.

according to this view. It is motivated by the articulatory bottleneck, which claims that communication proceeds remarkably quickly although humans can only produce a highly limited number of phonemes per second. It predicts that the upper-bound interpretation of a scalar implicature (i.e., *some, but not all*) precedes the lower-bound interpretation (*some and possibly all*) and thus the semantic (or lower bound) meaning of *some* should require more processing resources and take longer than processing of the pragmatic (or upper-bound) meaning of *some.*

Experimental evidence. The empirical picture on implicatures is complex, as they have been shown to be influenced by a variety of factors and have been investigated extensively with a variety of methods. Due to space limitations, we cannot discuss the experimental literature in detail here. An overview of the historical development of the experimental literature on implicatures can be found in Noveck (2018). Broadly speaking, implicature derivation has been shown to be more difficult than accessing literal meaning, supporting a literal-first view. Results diverge with regard to when effects occur. Some studies reported delayed processing of implicatures associated with *or* and *some* (e.g., Noveck & Posada, 2003; Bott & Noveck, 2004; Breheny et al., 2006; Chevallier et al., 2008; Huang & Snedeker, 2009, 2011; Bott et al., 2012). However, other experimental data, for example from eye-tracking (Grodner et al., 2010; Breheny et al., 2013; Foppolo & Marelli, 2017), suggest that implicatures are immediately available (see also Sedivy et al., 1999). Crucially for our purposes, differences in processing between presuppositions and scalar implicatures seem to confirm their different theoretical status (Bill et al., 2018).

3 The Experiment

Our experiment is a conceptual replication of Tiemann et al.'s (2011) Experiment 2. Like these authors, we employed self-paced reading to investigate at which point in time a presupposition verification with the context takes place. Tiemann et al. included five different presupposition triggers in their experiment (German *wieder*, Engl. *again*; *auch*, Engl. *also*; *aufhören*, Engl. *stop*; *wissen*, Engl. *know*; and definites in the shape of possessive noun phrases [German *sein/ihr*, Engl. *his/her*]), and compared sentences in contexts that either (i) verified the presupposition or (ii) falsified it. The language of investigation was German. Reading times at the positions of the trigger, the word following the trigger, and the evaluation word (the word where the content of the presupposition is known) were analyzed. The results supported the view that a validation process of the presupposition starts as soon as the presupposition is known, and before the end of the sentence. This validation process took longer in a falsifying context compared to a context that verified the presupposition. Furthermore, some studies suggested that different types of presupposition triggers may differ in processing (Domaneschi et al., 2014; Tiemann et al., 2015; Domaneschi & Di Paola, 2018). Against this background, separate analyses of presupposition triggers seem necessary, but Tiemann et al.'s (2011) analysis did not distinguish between different types of triggers, likely because of an insufficient number of data points per trigger.

Here, we focus on two presupposition triggers, the definite determiner *the* (German *der/die/das*) and its uniqueness-presupposition and *again* (German *wieder*). This focus allows for using a larger number of items per participant and condition and, as a result, meaningful analyses of differences between the two presupposition triggers. We additionally included scalar implicatures to contrast their processing with that of presuppositions. Sentences with scalar *some* (German *einige*) were used for the implicature condition, and sentences with *all* (German *alle)* were included as a control. They should not evoke an implicature as they are already most informative, and can only be either literally true or false.

3.1 General Approach and Procedure

For each sentence type (determiner, *again*, scalar implicature), we created 40 sets of experimental sentences, thus 120 sets in total. Each context sentence was paired with two test sentences in such a way that the context sentence either verified the presupposition/scalar implicature of the test sentence or falsified it. If context (A) verified the content of the presupposition/scalar implicature, then context (B) falsified it and vice versa (see Table 1 for example items).

To facilitate the comparison of the different regions of interest we refer to the words of interest via the region they appear in: P1 is the "trigger word" (scalar term for scalar implicatures and lexical trigger for presuppositions; underlined in Table 1). P2 is the critical word, that is, the point when the complete content of inferences was known, and P3 is the final word. The additions +1/+2 refer to the words following P1, P2, and P3, respectively (see also Table 2 in the Appendix).

Table 1. Example items with the two context variations in the verifying and falsifying condition

Example item: determiner					
(A)	Manuel hat ein Ticket für ein Baseballspiel gekauft.		(B)	Manuel hat mehrere Tickets für ein Baseballspiel gekauft.	
	Manuel bought a ticket for a baseball match.			*Manuel bought several tickets for a baseball match.*	
	1	Manuel holt das Ticket und freut sich.		3	Manuel holt die Tickets und freut sich.
		Manuel collects the ticket and he is happy.			*Manuel collects the tickets and he is happy.*
	2	Manuel holt die Tickets und freut sich.		4	Manuel holt das Ticket und freut sich.
		Manuel collects the tickets and he is happy.			*Manuel collects the tickets and he is happy.*
Example item: again					
(A)	Lukas hat schon oft Pizza bestellt.		(B)	Lukas hat noch nie Pizza bestellt.	
	Lukas has often ordered pizza before.			*Lukas has never ordered pizza before.*	
	5	Heute hat Lukas wieder Pizza bestellt und wartet freudig.		7	Heute hat Lukas wieder keine Pizza bestellt und hat nichts zu essen.
		Today, Lukas ordered pizza again and waits happily.			*Today, Lukas did not order pizza again and has nothing to eat.*
	6	Heute hat Lukas wieder keine Pizza bestellt und hat nichts zu essen.		4	Heute hat Lukas wieder Pizza bestellt und wartet freudig.
		Today, Lukas did not order pizza again and has nothing to eat.			*Today, Lukas ordered pizza again and waits happily.*
Example item: SI					
(A)	Zwei von vier Schrauben sind kaputt.		(B)	Vier von vier Schrauben sind kaputt.	
	Two of four screws are broken.			*Four of the four screws are broken.*	

	9	Weil einige Schrauben kaputt sind müssen neue gekauft werden.		11	Weil alle Schrauben kaputt sind müssen neue gekauft werden.
		Because some of the screws are broken they have to buy new ones.			*Because all of the screws are broken they have to buy new ones.*
	10	Weil alle Schrauben kaputt sind müssen neue gekauft werden.		12	Weil einige Schrauben kaputt sind müssen neue gekauft werden.
		Because all of the screws are broken they have to buy new ones.			*Because some of the screws are broken they have to buy new ones.*

3.2 Purpose of the Present Study and Hypotheses

The present study pursued four major goals. First, we aimed to replicate the results of Tiemann et al.'s (2011) Experiment 2. Like these authors, we employed a self-paced reading experiment to investigate at which point in time a presupposition verification with the context takes place. We expect that falsification of a presupposition causes more processing difficulties than its verification, being reflected by longer reading times in the falsification than in the verification condition (= Hypothesis 1, H1). These difficulties are expected in different regions depending on whether the process is semantic (early effects) or pragmatic (late effects).

Second, we aimed at comparing processing of Class 1 with the Class 2 presupposition triggers and Hypothesis 2 (H2) states to observe differences in processing of the two presupposition triggers. For example, the effect of context condition might be larger for definite determiners than for *again* because in the latter sentences, the presupposition can be ignored without making the sentence entirely senseless, that is, assertion and presupposition are not dependent on each other.

Third, we explored processing of scalar implicatures in comparison to presuppositions. For scalar implicatures, processing difficulties are expected (H3) and, depending on the theoretical approach, these difficulties appear in different conditions: we expect longer reading times in the falsifying condition according to the Default view, but in the verifying condition according to the literal-first view. Furthermore, we expect differences between presuppositions and scalar implicatures. Hypothesis 4 (H4) reflects our expectation of early effects starting on the trigger for the presuppositions, while for scalar implicatures slower and/or delayed processing is expected.

The main difference between the present study and earlier ones is that we focused on a subset of triggers to be able to gather more data points for each trigger (thus a conceptual replication). This allows for more precise parameter estimations and, thus, separate analyses. In addition, we included scalar implicatures to allow a comparison between their processing and processing of presuppositions.

4 Method

4.1 Participants

The intended sample size in this experiment was $n = 48$ native speakers of German. Data were collected from 52 participants from the Tübingen (Germany) area. Two participants were excluded because of 30 % or more errors in the final comprehension questions (see below). Another participant was excluded because German was not his/her mother tongue, and one participant was excluded due to technical problems during the experiment (final sample: mean age = 23.1 years, 39 females, 9 males). Participants signed informed consent prior to data collection and were paid 8 € or received course credit for participation.

4.2 Apparatus and Stimuli

Stimulus presentation and response collection were controlled by a standard PC connected to a 17-inch CRT monitor. A trial started with a context sentence presented as a whole in the upper half of the screen. After reading the context sentence, participants requested the test sentence with a button press of the right index finger on an external response button placed to the right of them. The test sentence was presented word-by-word in a self-paced reading manner, that is, the letters of the test sentences' words were first substituted by underscores as placeholders. Pressing the external key revealed the first word. To continue reading, participants had to press the button again to reveal the next word while the previous word disappeared and was again substituted with underscores. For *again* sentences, we presented *again not* (*wieder keine*) together to facilitate a comparison of the respective regions. Admittedly, this was the only time two words were presented together, but we did so to keep the position of the critical word parallel for the positive and negative polarity version of the *again* sentences. As we present reading times per letter, this decision is not problematic for our predictions and results. Effects of word length did not play a role because we analyzed reading time per letter (see Tiemann et al., 2011; Schneider et al., 2020b). After participants finished reading the test sentence, they had to rate its appropriateness according to the presented context via the number keys 1-4 on a standard QWERTZ keyboard ranging from very unnatural (1) to very natural (4). Participants were asked to pay attention to the content because after the experiment they had to answer ten yes/no comprehension questions (on the basis of which two participants were excluded).

Participants started with reading instructions. This was followed by a short practice block of 24 trials, four of each sentence type in both conditions. The order of these practice trials was determined randomly, but then kept constant for all participants. Subsequently, the 240 test trials were administered in three blocks of 80 trials each. The trials were presented in random order with the restriction that sentences of the same item did not appear in different conditions in direct succession. All participants were tested individually in a single session of about 45 minutes.

4.3 Design and Analyses

The independent variables of interest were (1) context condition (verifying, falsifying) and (2) sentence type (determiner, *again*, *some*, *all*). Mean acceptability ratings were submitted to a 2 × 4 Analysis of Variance (ANOVA) with context condition and sentence type as repeated measures.

There are three regions of interest (see also Table 2 in the Appendix): the position of the presupposition trigger or the scalar term (Position 1; P1), the word of evaluation, which is the word where the content of the presupposition or the scalar implicature could be evaluated (Position 2; P2), and the final word of the sentence (Position 3, P3). To be able to uncover spillover effects, we additionally analyzed one word following the trigger/scalar term (P1+1), one word following the evaluation word (P2+1), and two words following the evaluation word (P2+2). In case of the determiner, the word following the presupposition trigger (P1+1) is already the evaluation word (P2) (except for two items which were excluded from the analysis). Therefore, we used the same data in the ANOVA for both positions. Trials in which reading times deviated more than 2.5 *SD*s from the respective design cell in any of the analyzed positions (calculated separately for each participant) were excluded as outliers (15.0 % of the trials). For each region, mean reading times were submitted to the same ANOVA as acceptability ratings. In case of a significant interaction, we analyzed the sentence types separately with follow-up ANOVAs with context condition as a repeated measure.

5 Results

5.1 Acceptability Rating

Results of the acceptability ratings are visualized in Figure 1. Verifying conditions were clearly rated better than falsifying conditions for all four sentence types, $F(1, 47) = 676.46$, $p < .001$, $\eta_p^2 = .94$. The main effect of sentence type was also significant, $F(3, 141) = 27.74$, $p < .001$, $\eta_p^2 = .37$, reflecting the slight differences between the sentence types. The high ratings in the verifying condition suggest that participants perceive them as appropriate. In contrast, the falsifying condition received low ratings which indicates that inappropriateness was detected. The interaction was also significant, $F(3, 141) = 31.28$, $p < .001$, $\eta_p^2 = .40$, but the effect of context condition was significant for all sentence types, all ps $< .001$.

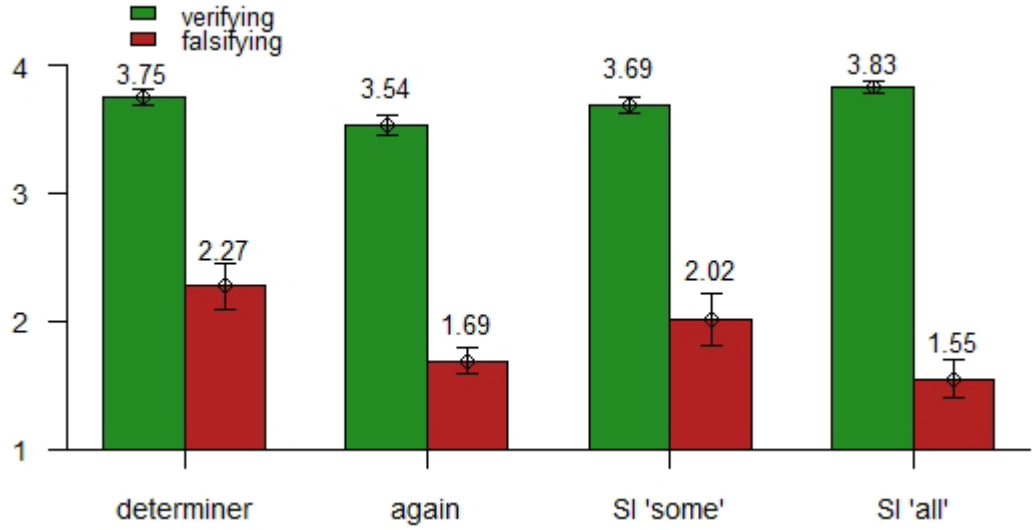

Figure 1. Mean acceptability ratings as a function of context condition and sentence type. Error bars are 95 % confidence intervals of the mean (SI = scalar implicature)

5.2 Reading Times

Reading times per letter are visualized in Figure 2 for all sentence types. Details on inferential statistics are summarized in Table 3 in the Appendix. In short, the ANOVAs revealed significant differences between the four sentence types for all analyzed positions, indicating that different processing difficulties were evoked by these sentence types.

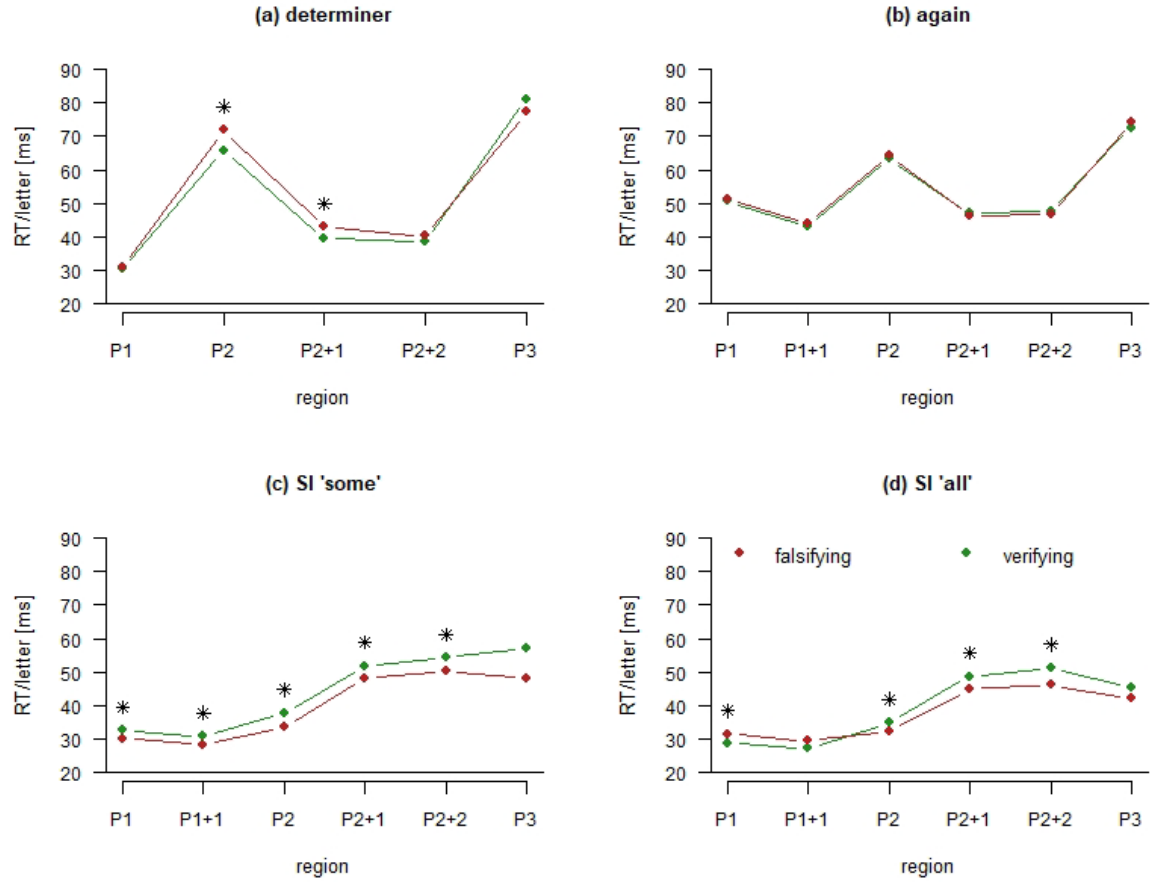

Figure 2. Reading times (RT; in milliseconds per letter) separately for the four sentence types determiner in (a), *again* in (b), implicatures with *some* in (c) and for true/false assertion with *all* in (d) for the respective regions. The asterisk marks significant differences between the sentence conditions (*P1* reflecting the trigger/scalar term, *P1+1* the word following the trigger/scalar term, *P2* the word of evaluation where the presupposition/scalar implicature can be evaluated, *P2+1* one word following the word of evaluation, *P2+2* two words following the word of evaluation, *P3* the final word of the sentence, SI = scalar implicature)

For the presupposition trigger/scalar term (P1), the main effect of context condition was not significant. However, we observed a significant interaction of context condition and sentence type. Context condition only had a significant impact for *some* sentences (with longer reading times in the verifying context condition) and *all* sentences (with longer reading times in the falsifying context condition), but not for the two presupposition triggers.

For the word following the presupposition trigger/scalar term (P1+1), additionally the main effect of context condition and the interaction were significant. The effect of context condition was only significant for determiner sentences (longer reading times in falsifying contexts) and for *some* sentences (longer reading times in verifying contexts).

At the evaluation word (P2), the main effect of context condition was not significant, but the interaction was. Context condition had an effect for determiners (longer reading times in falsifying contexts) and for *some* and *all* sentences (longer reading times in verifying contexts).

For the word following the evaluation word (P2+1) the main effect of context condition and the interaction were significant as well. Effects of context condition were observed for determiners (with longer reading times in the falsifying condition) and for *some* and *all* sentences (with longer reading times in the verifying condition). For *all* sentences, this could reflect the point when the sentence stopped making sense to participants (i.e., when they know the assertion is false in the context).

For the next word of interest (P2+2), the main effects of context condition and the interaction were significant. Context condition only had a significant impact on *some* and *all* sentences (with longer reading times in the verifying condition).

For the final word (P3), neither the main effect of context condition nor the interaction were significant. Descriptively, reading times for *again* were longer in the falsifying than in the verifying condition, while the opposite was true for the other three sentence types (with a slightly larger difference for *some* and *all* sentences as compared to determiner sentences).

6 Discussion

The results of the acceptability ratings replicate those of Tiemann et al. (2011) by and large. The ratings for all sentence types are high in the verifying condition and low in the falsifying condition. Accordingly, when the context does not verify the presupposition/scalar implicatures, it is perceived as inappropriate or unnatural. The scalar implicature *all* sentences can be regarded as a control because in these sentences verifying/falsifying means literally true vs. false. Thus, without requiring pragmatic enrichment, the verifying condition was perceived as natural.

Reading time results are only partly in line with Tiemann et al.'s (2011) observations. For the trigger position, we observed the same pattern as in the original study, namely that reading times did not differ between the two context conditions. Regarding performance on the evaluation word, our observations differed from Tiemann et al.'s results. For determiner sentences, we observe longer reading times in the falsifying condition than in the verifying condition. This result suggests an immediate verification process of the presupposition as soon as the content of the presupposition is fully known and thus supports the semantic approach to presuppositions (H1). If the content of the presupposition is not supported by the context, this process fails, leading to processing difficulties which become reflected in longer reading times at the evaluation word (but not at the end of the sentence). Thus, the data provide evidence for the semantic approach to presuppositions triggered by definite determiners and are not in line with a strictly pragmatic approach.[3] However, we do not see this difference for *again* sentences. This difference highlights the necessity to analyze different presupposition triggers separately, just as is suggested by the classifications of Tiemann et al. (2015) and Glanzberg (2005) (see also Domaneschi et al., 2014, 2018; Jouravlev et al., 2016; Domaneschi & Di Paola, 2018). As the processing of the two triggers under investigation here differs, the data are in line with H2.

For scalar implicatures, there is a significant difference between the two context conditions already at the position of the scalar term. The longer reading times in the verifying condition for the *some* sentences support the idea that scalar implicatures are only calculated in context conditions that support the implicature (see also Breheny et al., 2006; Hartshore, 2015). Regarding H3, the data provide evidence for the literal-first view: sentences with the weaker scalar term led to processing difficulties in the verifying condition, when the implicature had to be calculated, not in the falsifying condition, where no implicature needed to be derived. We also find, contrary to previous work by Noveck & Posada (2003) and Huang & Snedeker (2011), that this process starts quite early, and thus is in line with work by Grodner et al. (2010), Breheny et al. (2013), and Foppolo & Marelli (2017) who suggested that implicatures are immediately available. This speaks in favor of enrichment taking time.

Last, we can see that processing of scalar implicatures differs from processing presuppositions (H4) because for scalar implicatures the effect of context condition persists longer, until the end of the sentence. Thus, in contrast to presuppositions, the evaluation process of scalar implicatures appears to be a long-lasting process.

[3] At least one that does not assume local contexts to play a role for pragmatic processing/context integration.

There are two unexpected and interesting results. First, the reversed effect of context condition for the *all* sentences requires attention. Tentatively, we suggest the following explanation: the longer reading times for the falsifying context at the beginning of the sentence suggest that false assertions are detected rapidly by participants. This leads to processing difficulties at the scalar term and at the following word. At that point, participants realized that the sentence stopped making sense and consequently they did not process the assertion in the falsifying condition properly. This could then explain the shorter reading times in the falsifying condition in the later parts of the sentence.

Second, we did not observe a strong effect of context for the trigger *again,* contrary to previous findings. This may be either due to the material we used, which is possibly too unspecific with regard to the truth of the presuppositions. It may also be due to the very complex interaction of *again* with negation, which makes verification and falsification more complex (see Schwarz & Tiemann, 2017 for more discussion). We have to leave it to further research to address these issues in more detail.

7 Conclusion

The present study set out to replicate and extend previous results regarding immediate verification processes of presuppositions in contrast to scalar implicatures. Two main conclusions can be drawn based on the experiment we reported: first, for presuppositions triggered by definite determiners, a verification process is immediately started as soon as the content of the presupposition is known. This is in line with a semantic approach according to which the presupposition is encoded in the lexical entry of the presupposition trigger. This process evokes longer reading times in case of failure than in a successful verification. This pattern was not observed for presuppositions triggered by *again* though; a result highlighting the need of separate analyses of different presupposition triggers. Second, the data reveals that processing of presuppositions and scalar implicatures differs. Scalar implicatures are not automatically evaluated. Only in case of a verification by the context, the implicature is computed. Otherwise, the implicated meaning is not evaluated. The data revealed that the evaluation process of the implicature starts early already at the scalar term, but persists till the end of the sentence. This is in line with a view predicting the processing of scalar implicatures to require more effort.

References

Abrusán, M. (2011). Predicting the presuppositions of soft triggers. *Linguistics and Philosophy, 34*, 491-535.

Abusch, D. (2002). Lexical alternatives as a source of pragmatic presupposition. In B. Jackson (Ed.), *Proceedings of Semantics and Linguistic Theory 12* (pp. 1-19). San Diego, CA: University of California and San Diego State University.

Abusch, D. (2010). Presupposition triggering from alternatives. *Journal of Semantics, 27*(1), 37-80.

Bade, N., & Schwarz, F. (2019). An experimental investigation of antipresuppositions. In A. Creemers & C. Richter (Eds.), *Proceedings of PLC 42* (pp. 29-40). Philadelphia, PA: University of Pennsylvania Working Papers in Linguistics.

Beaver, D., & Geurts, B. (2012). Presupposition. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 3, pp. 2432-2459). Berlin: Mouton de Gruyter.

Bill, C., Romoli, J., & Schwarz, F. (2018). Processing presuppositions and implicatures: Similarities and differences. *Frontiers in Communication, 3*, 44.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language, 66*(1), 123-142.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*(3), 437-457.

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes, 28*(4), 443-467.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition, 100*(3), 434-463.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language, 98*(2), 159-168.

Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology, 61*(11), 1741-1760.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Belletti Structures and Beyond* (pp. 39-103). Oxford: Oxford University Press.

Chierchia, G., (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry, 37*(4), 535-590.

Domaneschi, F., Canal, P., Masia, V., Lombardi Vallauri, E., & Bambini, V. (2018). N400 and P600 modulation in presupposition accommodation: The effect of different trigger types. *Journal of Neurolinguistics, 45,* 13–35.

Domaneschi, F., Carrea, E., Penco, C., & Greco, A. (2014). The cognitive load of presupposition triggers: Mandatory and optional repairs in presupposition failure. *Language, Cognition and Neuroscience, 29*(1), 136-146.

Domaneschi, F., & Di Paola, S. (2018). The processing costs of presupposition accommodation. *Journal of Psycholinguistic Research, 47*(3), 483-503.

Foppolo, F., & Marelli, M. (2017). No delay for some inferences. *Journal of Semantics, 34*(4), 659-681.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland & P. Stateva (Eds.), *Presupposition and Implicature in Compositional Semantics* (pp. 71-120). Basingstoke, Hampshire: Palgrave Macmillan.

Glanzberg, M. (2005). Presuppositions, truth values and expressing propositions. In G. Preyer & G. Peter (Eds.), *Contextualism in Philosophy: Knowledge, Meaning, and Truth* (pp. 349-396). Oxford: Oxford University Press.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). Some, and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition, 116*(1), 42-55.

Hartshorne, J. K., Snedeker, J., Liem Azar, S. Y.-M., & Kim, A. E. (2015). The neural computation of scalar implicature. *Language, Cognition and Neuroscience, 30*(5), 620-634.

Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar*. New York, NY: Wiley-Blackwell.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology, 58*(3), 376-415.

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes, 26*(8), 1161-1172.

Jouravlev, O., Stearns, L., Bergen, L., Eddy, M., Gibson, E., & Fedorenko, E. (2016). Processing temporal presuppositions: An event-related potential study. *Language, Cognition and Neuroscience, 31*(10), 1245-1256.

Kirsten, M., Tiemann, S., Seibold, V. C., Hertrich, I., Beck, S., & Rolke, B. (2014). When the polar bear encounters many polar bears: Event-related potential context effects evoked by uniqueness failure. *Language, Cognition and Neuroscience, 29*(9), 1147-1162.

Levinson, S. (2000). *Presumptive Meanings*. Cambridge, UK: Cambridge University Press.

Magri, G. (2010). *A Theory of Individual-Level Predicates Based on Blind Mandatory Scalar Implicatures*. Dissertation, Massachusetts Institute of Technology.

Noveck, I. (2018). How logical terms can be enriched: Exposing semantic-pragmatic divergences. In *Experimental Pragmatics: The Making of a Cognitive Science* (pp.78-101). Cambridge, UK: Cambridge University Press.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language, 85*(2), 203-210.

Romoli, J. (2011). The presuppositions of soft triggers aren't presuppositions. In N. Ashton, A. Chereches & D. Lutz (Eds.), *Proceedings of Semantics and Linguistic Theory 21* (pp. 236-256). New Brunswick, NJ: Rutgers University.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy, 27*, 367-391.

Schlenker, P. (2011). DRT with local contexts. *Natural Language Semantics, 19*, 373-392.

Schlenker, P. (2012). Maximize presupposition and Gricean reasoning. *Natural Language Semantics, 20*, 391–429.

Schneider, C., Bade, N., Franke, M., & Janczyk, M. (2020a). Presuppositions of determiners are immediately used to disambiguate utterance meaning. *Psychological Research, 44*, 1931-1946.

Schneider, C., Bade, N., & Janczyk, M. (2020b). Is immediate processing of presupposition triggers automatic or capacity-limited? *Journal of Psycholinguistic Research, 49*, 247-273.

Schneider, C., Schonard, C., Franke, M., Jäger, G., & Janczyk, M. (2019). Pragmatic processing: An investigation of the (anti-)presupposition of determiners using mouse-tracking. *Cognition, 193*, 104024.

Schwarz, F. (2007). Processing presupposed content. *Journal of Semantics, 24*(4), 373-416.

Schwarz, F. (2016). Experimental work in presupposition and presupposition projection. *Annual Review of Linguistics, 2*, 273-292.

Schwarz, F., & Tiemann, S. (2017). Presupposition projection in online processing. *Journal of Semantics, 34*(1), 61-106.

Sedivy, J., Tanenhaus, M., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretations through contextual representation. *Cognition, 71*(2), 109-147.

Simons, M. (2001). On the conversational basis of some presuppositions. In R. Hastings, B. Jackson & Z. Zvolenszky (Eds.), *Proceedings of Semantics and Linguistics Theory 11* (pp. 431-448). New York, NY: New York University.

Simons, M., Tonhauser, J., Beaver, D., & Roberts, C. (2010). What projects and why. In N. Li & D. Lutz (Eds.), *Proceedings of Semantic and Linguistic Theory 20* (pp. 309-327). Vancouver: CLC Publications.

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic, 2*(4), 447-457.

Strawson, P. (1952). *Introduction to Logical Theory*. London, UK: Routledge.

Tiemann, S. (2014). *The Processing of wieder ('again') and other Presupposition Triggers*. Dissertation, University of Tübingen.

Tiemann, S., Kirsten, M., Beck, S., Hertrich, I., & Rolke, B. (2015). Presupposition processing and accommodation: An experiment on wieder ('again') and consequences for other triggers. In F. Schwarz (Ed.), *Experimental perspectives on presuppositions. Studies in theoretical psycholinguistics* (pp. 39-65). Cham: Springer.

Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., Knapp, J., & Beck, S. (2011). Psycholinguistic evidence for presuppositions: On-line and off-line data. In I. Reich, E. Horch & D. Pauly (Eds.), *Proceedings of Sinn & Bedeutung 15* (pp. 581-595). Saarbrücken: Saarland University Press.

Tonhauser, D. (2015). Are 'informative presuppositions' presuppositions? *Language and Linguistic Compass, 9/2*, 77-101.

van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*(2), 147-182.

van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*(2), 235-248.

von Fintel, K. (2004). Would you believe it? The King of France is back! (Presuppositions and truth-value intuitions). In M. Reimer & A. Bezuidenhout (Eds.), *Descriptions and Beyond* (pp. 315-341). Oxford: Oxford University Press.

Appendix

Table 2. Regions of interest for the respective conditions, in bold significant effects

	P1: presupposition trigger/ scalar term	**P1+1: post-trigger**	**P2: evaluation**	**P2+1: post-evaluation**	**P2+2: post-evaluation 2**	**P3: final word**
Context	$F(1,47)$ = 0.71, p = .405, η_p^2 < .01	**$F(1,47)$ = 6.91, p = .012, η_p^2 = .13**	$F(1,47)$ = 0.00, p = .980, η_p^2 < .01	$F(1,47)$ = 4.59, p = .037, η_p^2 = .09	**$F(1,47)$ = 9.44, p = .004, η_p^2 = .17**	$F(1,47)$ = 3.69, p = .061, η_p^2 = .07
Sentence type	**$F(3,141)$ = 79.51, p < .001, η_p^2 = .63, ε = .69**	**$F(3,141)$ = 180.70, p < .001, η_p^2 = .79, ε = .58**	**$F(3,141)$ = 174.55, p < .001, η_p^2 = .79, ε = .47**	**$F(3,141)$ = 30.74, p < .001, η_p^2 = .40, ε = .75**	**$F(3,141)$ = 52.00, p < .001, η_p^2 = .53, ε = .80**	**$F(3,141)$ = 42.91, p < .001, η_p^2 = .48, ε = .62**
Interaction	**$F(3, 141)$ = 4.60, p = .012, η_p^2 = .09**	**$F(3,141)$ = 5.35, p = .004, η_p^2 = .10, ε = .76**	**$F(3,141)$ = 10.36, p < .01, η_p^2 = .18, ε = .55**	**$F(3,141)$ = 9.66, p < .001, η_p^2 = .17, ε = .84**	**$F(3,141)$ = 6.69, p = .001, η_p^2 = .12, ε = .75**	$F(3,141)$ = 1.93, p = .128, η_p^2 = .04
determiner	$F(1,47)$ = 0.32, p = .576, η_p^2 = .01	**$F(1,47)$ = 6.99, p = .011, η_p^2 = .13**		**$F(1,47)$ = 13.94, p < .001, η_p^2 = .23**	$F(1,47)$ = 3.13, p = .083, η_p^2 = .06	
again	$F(1,47)$ = 0.34, p = .562, η_p^2 = .01	$F(1,47)$ = 0.55, p = .464, η_p^2 = .01	$F(1,47)$ = 0.98, p = .328, η_p^2 = .02	$F(1,47)$ = 1.21, p = .277, η_p^2 = .03	$F(1,47)$ = 2.74, p = .105, η_p^2 = .06	
SI *some*	**$F(1,47)$ = 5.58, p = .022, η_p^2 = .11**	**$F(1,47)$ = 4.96, p = .031, η_p^2 = .10**	**$F(1,47)$ = 17.50, p < .001, η_p^2 = .27**	**$F(1,47)$ = 6.77, p = .012, η_p^2 = .13**	**$F(1,47)$ = 5.50, p = .023, η_p^2 = .10**	
SI *all*	**$F(1,47)$ = 5.55, p = .023, η_p^2 = .11**	$F(1,47)$ = 3.69, p = .061, η_p^2 = .07	**$F(1,47)$ = 7.79, p = .008, η_p^2 = .14**	**$F(1,47)$ = 9.55, p = .003, η_p^2 = .17**	**$F(1,47)$ = 13.44, p = .001, η_p^2 = .22**	

Note: SI = scalar implicature

Table 3. Inferential statistics for Experiment 1. The first rows are the statistics for the 4 × 2 ANOVA for each region. In case of a significant interaction, separate ANOVAs with context condition as a repeated measure were run.

				P1	P1 +1		P2	P2+1	P2+2			P3
determiner		Manuel	holt	**das/die**			Ticket/s	und	freut			sich.

again		Heute	hat	Lukas	**wieder (keine)**	Pizza		bestellt	und	hat	nichts	zu	essen.
SI	***all***			Weil	**einige**	Schrauben	kaputt	sind	müssen	neue	gekauft		werden.
	some			Weil	**alle**	Schrauben	kaputt	sind	müssen	neue	gekauft		werden.

Note: P1 = presupposition trigger/scalar term; P2 = word of evaluation; P3 = final word; SI = scalar implicature