Language and Cognition **9** (2017), 603–636. doi:10.1017/langcog.2016.30
© UK Cognitive Linguistics Association, 2016

Are all the triangles blue? – ERP evidence for the incremental processing of German quantifier restriction*

PETRA AUGURZKY
Department of General Linguistics, University of Tübingen, and Sonderforschungsbereich 833, University of Tübingen

OLIVER BOTT
Sonderforschungsbereich 833, University of Tübingen

WOLFGANG STERNEFELD
Department of General Linguistics, University of Tübingen

AND

ROLF ULRICH
Department of Psychology, University of Tübingen

(*Received* 07 *June* 2016 – *Revised* 30 *August* 2016 – *Accepted* 06 *September* 2016 – *First published online* 27 *October* 2016)

ABSTRACT

The present ERP study investigates the neural correlates of pictorial context effects on compositional-semantic processing. We examined whether the incremental processing of questions involving quantifier restriction is modulated by the reliability of pictorial information. Contexts either allowed for an unambiguous meaning evaluation at an early sentential position or were ambiguous with respect to whether a further restrictive cue could trigger later meaning revisions. Attention was either guided towards (Experiment 1) or away from (Experiment 2) the picture–question mapping. In both experiments, negative answers elicited a broadly distributed negativity opposed to affirmative answers as soon as an unambiguous truth evaluation was possible. In the presence of ambiguous context information, the truth evaluation initially remained

[*] The current research was supported by the SFB 833 of the Deutsche Forschungsgesellschaft, Project *B1*. We would like to thank Fabian Schlotterbeck and Robin Hörnig for valuable advice, Nadine Balbach and Helena Schütze for help with the data acquisition, and Jochen Saile for programming advice. Address for correspondence: Petra Augurzky. e-mail: petra.augurzky@uni-tuebingen.de

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

underspecified, as an early commitment would have resulted in the risk of a semantic reanalysis. The negativity was followed by a late positivity in Experiment 1, but not in Experiment 2, suggesting that attention towards the mismatch affected semantic processing, but only at a later time window. The current results are consistent with the notion that an incremental meaning evaluation is dependent on the reliability of contextual information.

KEYWORDS: language comprehension, compositional-semantic processing, quantifier restriction, context effects, event-related potentials, N400, late positivity.

1. Introduction

Though sentence meaning is computed rapidly, contextual information may contribute to the semantic evaluation of an utterance, thus leading to processing difficulties when the two domains are not in accord. Therefore, the ultimate interpretation of a sentence is not only determined by the meanings of its words, but also by extra-sentential cues. For instance, sentences containing quantifiers, as in (1), may be interpreted differently depending on the location where they are uttered: when uttered in Germany, the sentence will likely be judged false, while it will be judged true when uttered in a busy Manhattan street.

(1) All cabs are yellow.

Quantifiers like *all* are grammatical determiners expressing abstract quantity information (Barwise & Cooper, 1981), and are generally taken to denote relations between two sets, their restrictor argument and their nuclear scope. In the above example, the restrictor set is a set of relevant cabs and the scope set is the set of yellow things, or alternatively, equivalent to the set of yellow things that are cabs (e.g., Peters & Westerståhl, 2006). Thus, one specific semantic property of quantifiers is that they trigger domain restriction over their nominal arguments (Westerståhl, 1985). In (1), the quantifier *all* is automatically restricted to a contextually relevant set of cabs instead of taking its argument to refer to all existing cabs in the world. For arriving at a final truth evaluation, contextual information thus needs to be combined with sentential meaning during some stage of processing. To date, relatively little is known about such semantic restriction processes during sentence comprehension.

The present event-related potentials (ERP) study investigates quantificational restriction in the presence of pictorial context information. For examining context effects during on-line semantic comprehension, electro-physiological

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

measures are particularly well suited (Hagoort & van Berkum, 2007). Event-related brain potentials provide a high temporal resolution for investigating the incremental nature of compositional processes, and are sensitive to distinct aspects of meaning. In the ERP literature, a majority of studies on context effects focuses on semantic word integration. In these studies, facilitating effects of contextual information were shown (Hagoort, Hald, Bastiaansen, & Petersson, 2004), which are reflected by the N400 component, a negative-going deflection with a centro-parietal maximum between 200 and 600 ms post stimulus onset. The amplitude of the N400 is gradually affected by the goodness of contextual fit, with the smallest amplitude for the best fit (Kutas & Hillyard, 1984; Kutas, Van Petten, & Kluender, 2006). Besides intra-sentential information, extra-sentential cues such as world knowledge, speaker identity, and pragmatic inferencing affect the N400 (Filik & Leuthold, 2008; Kutas & Federmeier, 2011; Nieuwland, Ditman, & Kuperberg, 2010; van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). These findings are compatible with an interpretation of the N400 as representing the semantic fit between any kind of contextually relevant information and the actually processed lexical item. Moreover, it has been shown that the N400 is modulated by particularly salient contextual information, which may elicit expectancies regarding the lexical–semantic properties of upcoming words (e.g., DeLong, Urbach, & Kutas, 2005; Ito, Corley, Pickering, Martin, & Nieuwland, 2016; Otten, Nieuwland, & van Berkum, 2007).

Rapid context effects were also reported in another domain of semantic processing, namely the comprehension of verb action information. Using a picture–sentence verification paradigm, Knoeferle, Urbach, and Kutas (2011) investigated German equivalents of sentences like *The gymnast pushes / applauds the journalist* that were preceded by pictures that correctly or incorrectly depicted the scenes described by the sentences. At the position of the verb, an increased N400 signalled a mismatch between the picture and the actually presented sentential information (see also Vissers, Kolk, van de Meerendonk, & Chwilla, 2008).

In contrast to these early effects, studies on the temporal availability of quantifier meaning suggest that at least some aspects of semantic composition are contextually constrained in a delayed manner. So-far investigated quantificational properties include domain restriction (Frazier, Clifton, Rayner, Deevy, Koh, & Bader, 2005; Kaan, Dallas, & Barkley, 2007), scope ambiguities (for a recent overview, see Bott & Schlotterbeck, 2015), monotonicity differences (Deschamps, Agmon, Loewenstein, & Grodzinsky, 2015), scalar implicatures (Hartshorne, Snedeker, Azar, & Kim, 2015; Huang & Snedeker, 2009; Hunt, Politzer-Ahles, Gibson, Minai, & Fiorentino, 2013, Politzer-Ahles, Fiorentino, Jiang, & Zhou, 2013), and quantificational consequences on reference resolution (see Nouwen, 2010, for review) or world knowledge

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

evaluations (Kounious & Holcomb, 1992; Nieuwland, 2016; Urbach & Kutas, 2010; Urbach, DeLong, & Kutas, 2015). Though it is uncontroversial that quantificational properties affect the overall interpretation of a sentence, it is still an unresolved issue whether quantifier meaning 'incrementally' determines the truth evaluation at each incoming word.

In fact, at least on the single-sentence level, there are cases in which quantificational information does not appear to contribute to an immediate meaning assignment, but is only evaluated at the end of the sentence. In particular, the processing of quantifier scope has been shown to give rise to non-incremental semantic processing effects. For instance, in an eye-tracking during reading experiment, Bott and Schlotterbeck (2015) examined the on-line effects of scope inversion in multiply quantified sentences. In sentences such as *Jeden dieser/seiner Schüler hat genau ein Lehrer voller Wohlwollen gelobt* 'each of-these / of-his pupils has exactly one teacher praised', processing difficulties for scope-inverted readings were delayed to the clause-final verb, though a local truth evaluation could have already been computed at the position of the second quantified NP 'exactly one teacher'. This finding indicates that a fully-fledged interpretation of quantifier scope inversion may require the exhaustive processing of the whole sentence (see also Dwivedi, Phillips, Einagel, & Baum, 2010, who did not find local N400 differences between linear and inverted scope; but see Dotlačil & Brasoveanu, 2015, for an opposing view).

To date, only a few studies have considered the incremental effects of contextual properties on semantic aspects of quantifier processing.[1] With regard to quantifier restriction, a study by Kaan et al. (2007) investigated extra-sentential context effects on the processing of quantified numerals as in *Twelve/four flowers were put in a vase. Six had a broken stem*, in which the meaning of the numeral *six* is interpreted with respect to the preceding subject NP. Following *twelve flowers*, the numeral could either refer to six of the twelve previously mentioned flowers or, alternatively, to a different reference set containing flowers that were not mentioned before. Whereas behavioural studies using comparable materials showed rather early effects (Frazier et al., 2005; Wjinen & Kaan, 2006), ERPs from the onset of the numeral *six* revealed that an impossible subset reading did not affect quantifier

[1] It has to be noted that several studies investigate context effects on pragmatic aspects of quantifiers such as *some* (see Spychalska, Kontinen, & Werning, 2016, for a recent review). Whereas the meaning of the quantifier *all* can be derived by logical-semantic processing, the meaning of the quantifier *some* is often assumed to be constrained by pragmatic processes. For instance, the sentence *Some cabs are yellow* will be usually interpreted as *some but not all cabs are yellow* due to processes of pragmatic enrichment (e.g., Grice, 1975; Noveck & Sperber, 2007). It is beyond the scope of the present paper to give an overview of these studies.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

processing within the N400 time-window. Instead, a late positivity was observed 900–1500 ms post numeral onset, possibly reflecting the establishment of a new discourse referent (Burkhardt, 2006; see also Paterson, Filik, & Moxey, 2009, for a comparable view). The delayed brain response suggests that a complete meaning evaluation was computed non-incrementally, i.e., on the word following the quantifier. This observation is probably not surprising, given that alternative continuations like *Six women …* are still available on the numeral, and an unambiguous decision could only be made on the verb *had*. These findings thus suggest that quantifier restriction is processed incrementally as soon as the input properties allow for an unambiguous decision.

Recently, Politzer-Ahles et al. (2013) investigated pictorial context effects on the processing of the quantifiers *all* and *some*. For the present considerations, the results for the quantifier *all* are therefore particularly relevant. Politzer-Ahles et al. tested Mandarin Chinese versions of the sentence *In this picture, all of the girls are sitting on blankets suntanning*, preceded by pictures that did or did not match the sentential meaning. In a first ERP experiment, a visual picture–sentence verification task, ERPs from quantifier onset showed a reduced negativity (200–500 ms) when the pictorial information was not in accord with the meaning conveyed by the quantifier. According to the authors, the negativity is a form of the Nref component reflecting a decreased effort to bind the quantifier with an antecedent once the semantic inconsistency with the context has been recognized. In the second experiment, employing a picture–sentence consistency rating, sentences were presented auditorily. Contrary to the results from the first experiment, semantically false sentences elicited a sustained positivity (300–1000 ms post quantifier onset). According to the authors, the divergent effects could be due to experiment-inherent differences, such as presentation mode or task demands. Similarly to the study by Kaan et al. (2007), the study by Politzer-Ahles et al. did not investigate processing on the point of disambiguation. As the authors note, an unambiguous decision could only be made later in the sentence. Since the materials in the post-quantifier region were not controlled for frequency, word order, and length, post-hoc analyses were not possible due to differing disambiguation regions (2013, p. 145f.).

Indeed, the occurrence of ambiguity may be one of the reasons for a delayed processing of specific quantificational properties, such as their effects on readers' on-line evaluations of sentences related to world knowledge. For instance, as pointed out by Nieuwland (2016, p. 317), a truth evaluation of the sentence fragment *Few majors see*, though locally false, might be delayed until a later position where quantifier scope can be unambiguously calculated (e.g., *Few majors see ghosts*). The avoidance of such potential meaning shifts may thus lead to underspecified semantic processing (see also Villalta, 2003).

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

However, even with unambiguous scope, quantifiers like *few* do not appear to be interpreted in a fully incremental fashion (e.g., Urbach & Kutas, 2010). Given the immediate context effects on lexical–semantic processing, a reasonable assumption would be that contextual information is already able to trigger more immediate decisions at the earliest possible sentential position.

In a recent study, Urbach et al. (2015) tested this assumption by examining contextual effects on quantified sentences related to world knowledge. When presented in isolation, quantificational properties have been shown to contribute inconsistently (Urbach & Koutas, 2010) or not at all (Kounious & Holcomb, 1992) to the incremental processing of such sentences. For instance, Kounios and Holcomb presented class inclusion statements such as *all / some / no rubies are gems / spruces* in an ERP study. Though off-line truth-value judgements showed that participants had correctly interpreted these sentences, ERPs from the onset of the sentence-final word showed an N400 reduction for *gems* vs. *spruces* regardless of the quantificational meaning, thus indicating that quantifier interpretation did not incrementally contribute to meaning assignment (see also Fischler, Bloom, Childers, Roucos, & Perry, 1983, for comparable results in the field of negation processing). Moreover, Urbach and Kutas (2010) investigated the effects of positive and negative quantifiers like *most* and *few* on the processing of sentences involving typical and atypical sentence continuations (*Most / few tourists visit museums / mines in their vacation*). Whereas post-sentential plausibility ratings showed the expected cross-over interaction (i.e., reduced plausibility ratings for atypical continuations *(mines)* for *most*, and the opposite pattern for *few*), ERPs from the sentential object showed a general typicality effect on the N400, which was reduced for negative quantifiers. This absence of a full cross-over interaction in the ERPs led the authors conclude that quantifier interpretation proceeds partly under-specified, with delayed processing of negative quantifiers.

In order to test whether contextual effects may boost the incremental in-depth processing of negative quantifiers, the follow-up study by Urbach et al. (2015) presented comparable sentences that were preceded by sentential contexts (e.g., *The contractor was shocked to see that the construction crew was comprised completely of women. This was quite unusual considering that # most / few construction workers are male / female in this day and age.*) By increasing the predictability of atypical continuations for the negative quantifiers, contexts should thus enhance the probability of obtaining a cross-over interaction already locally, as indexed by the N400. Results from two ERP studies show that, principally, contextual information is indeed able to trigger a fully incremental interpretation of negative quantifiers. However, the expected N400 effect was only observed in the second experiment,

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

a sentence reading task, whereas in the first study, a plausibility rating task, a similar discrepancy between on-line and off-line effects as in previous studies was observed. Though the reasons for this task dependency are currently unclear, it can be noted that, at least principally, contextual information may boost earlier in-depth semantic processing of quantificational properties (see also Nieuwland & Kuperberg, 2008, for a comparable effect on the processing of negated sentences). Thus, contextual licensing is principally able to determine the sentential positions where semantic commitments are made. Finally, incremental N400 effects for negative quantifiers have also been reported in Nieuwland (2016), who employed the same factorial design as Urbach and Kutas (2010), but additionally controlled for the predictability of the critical word. For items with a high cloze probability of the critical word across quantifier conditions, a fully-fledged cross-over interaction was found, but for items with low cloze probability, effects were similar to Urbach and Kutas' study. Quantifier interpretation thus can be fully incremental if there is sufficiently strong guidance by world-knowledge for predicting how the quantified statement will continue (see also Freunberger & Nieuwland, 2016, for facilitating effects of prosodic information).

In sum, contextual information has been shown to rapidly affect lexical–semantic word integration in the N400 time-window, as well as the processing of verb action information in the N400 time-window. With respect to quantifier interpretation, contextual cues are able to incrementally affect compositional-semantic processing, but the current findings are far from offering a homogeneous picture. The few studies explicitly dealing with quantifier restriction show a discrepancy between behavioural and ERP studies, with apparently earlier effects found in behavioural tasks. In these studies, quantifier processing was not established independently of discourse processes, and the delayed ERP response might be partially related to the fact that an unambiguous response is available only at the post-quantificational position. Comparably, the results from Politzer-Ahles et al. (2013) do not provide information about quantifier processing at a fully disambiguated position. On the other hand, contextual information has been shown to boost the incremental processing of quantified sentences dealing with world knowledge. However, though contextual cues affected processes already in the N400 time-window, the dependence on a specific experimental task makes it currently unclear whether these effects can be generalized.

The present study was designed to shed further light on the incremental nature of contextual effects on the processing of quantificational meaning. In particular, we included well-defined disambiguation regions to observe quantifier restriction as the sentence meaning unfolds over time.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

2. The present study

In two experiments, we examined quantifier restriction in sentences containing the quantifier *all* preceded by pictorial contexts. In contrast to previous studies, we controlled the sentential positions where unambiguous decisions could be made and also added a new dimension to previous work by investigating potential semantic revision effects as in *All cabs are yellow that are shown in this picture of Manhattan*. As noted above, these sentences may be judged as false on the colour adjective. However, a following restriction could trigger potential meaning shifts from 'false' to 'true'. For instance, the extraposed relative clause *that are shown in this picture of Manhattan* further restricts the set of cabs to a proper subset reading. In our study, we abstracted away from potential lexico-semantic confounds and presented materials involving geometrical figures. In both studies, one of four pictorial contexts (A–D, Figure 1) preceded a question as in (2), which was related to the content of the picture.

(2) Sind alle Dreiecke blau, die innerhalb des Kreises sind?
Are all triangles blue, that inside-of the circle are?
'Are all the triangles blue that are inside of the circle?'

We examined whether participants incrementally use pictorial cues during the processing of quantified questions. In addition, we tested whether semantic commitments are delayed in the face of ambiguity (cf. Kaan et al., 2007). If contextual information constrains semantic processing independently of potential ambiguities, then a semantic commitment will already be made at the earliest possible position, i.e., the colour adjective. Alternatively, if a truth evaluation is delayed until unambiguous information is encountered, a semantic commitment will only be made at the preposition in the relative clause. Note that an early local commitment might be risky, as it would come with the cost of a semantic revision. For instance, if (2) were preceded by context A, an initial 'false' judgement on the adjective needed to be revised to a 'true' judgement on the preposition. Therefore, we also tested whether the parser is sensitive to contextual reliability.

In order to control for potential confounds by silent prosody (Fodor, 2002), we presented target *yes/no* questions instead of declarative clauses. German declaratives are prosodically realized by clause-final falling contours (Eckstein & Friederici, 2006), which are incompatible with a relative clause continuation (Augurzky, 2006). By contrast, the prosody of questions is highly comparable to a relative clause continuation (both have final rises) and should thus render the sentence ambiguous with respect to whether or not a restrictive cue will follow.

To sum up: the present study examines context effects on the timecourse of quantifier restriction at different question positions. We tested whether

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

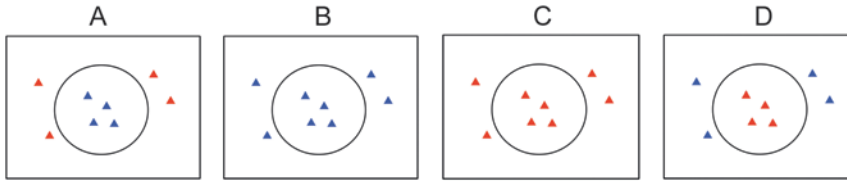


Fig. 1. Visual contexts preceding the target questions in Experiment 1.

contextual information constrains quantifier interpretation at the earliest possible position and whether potential meaning shifts occur at later positions. We examined whether ERP effects reflect early semantic integration processes within the N400 time-window or whether quantifier interpretation is generally delayed, as reflected by later positive components. To this end, we carried out two ERP studies. In our first study, we employed a picture–question verification task. In the second study, we used a probe detection task to test whether similar effects would be observed when attention was directed away from the picture–question mapping.

3. Experiment 1

In Experiment 1, a picture–question verification task, we examined whether context information incrementally constrains the semantic processing of questions involving quantificational restriction. To this end, one of four visual contexts (Figure 1) was presented at the beginning of each trial, and then a question containing the universal quantifier.

For examining restrictive processes during question processing, we compared ERPs at different sentential positions: first, from the onset of the colour adjective ('blue'), and second, from the onset of the preposition ('inside-of/outside-of'). The complete question materials are provided in Table 1. Contexts either consisted of simple (B,C) or complex (A,B) pictures. In each picture, geometrical objects like triangles were presented within and outside of a container form (e.g., a circle). In the simple pictures, all objects were of identical colour, whereas in the complex pictures, objects within and outside of the container form differed in colour.

Once the pictorial information has been processed, an answer to the following question is principally already available at the position of the colour adjective. If the parser incrementally makes a semantic commitment at this position, meaning shifts should be observed on the preposition following complex contexts (i.e., for A1 vs. A2 and D2 vs. D1). Given the currently indecisive results concerning the incrementality of context effects on quantificational processing, we intended to empirically discriminate between the following two competing hypotheses:

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

TABLE 1. *Design of the experiments*

Condition		Question	Values	Shift
A1	Context A, *inside-of*	Sind alle Dreiecke **blau**, die **innerhalb** des Kreises sind? 'Are all triangles blue, that inside-of the circle are?'	false → true	**yes**
A2	Context A, *outside-of*	Sind alle Dreiecke **blau**, die **außerhalb** des Kreises sind? 'Are all triangles blue, that outside-of the circle are?'	false → false	no
B1	Context B, *inside-of*	Sind alle Dreiecke **blau**, die **innerhalb** des Kreises sind? 'Are all triangles blue, that inside-of the circle are?'	true → true	no
B2	Context B, *outside-of*	Sind alle Dreiecke **blau**, die **außerhalb** des Kreises sind? 'Are all triangles blue, that outside-of the circle are?'	true → true	no
C1	Context C, *inside-of*	Sind alle Dreiecke **blau**, die **innerhalb** des Kreises sind? 'Are all triangles blue, that inside-of the circle are?'	false → false	no
C2	Context C, *outside-of*	Sind alle Dreiecke **blau**, die **außerhalb** des Kreises sind? 'Are all triangles blue, that outside-of the circle are?'	false → false	no
D1	Context D, *inside-of*	Sind alle Dreiecke **blau**, die **innerhalb** des Kreises sind? 'Are all triangles blue, that inside-of the circle are?'	false → false	no
D2	Context D, *outside-of*	Sind alle Dreiecke **blau**, die **außerhalb** des Kreises sind? 'Are all triangles blue, that outside-of the circle are?'	false → true	**yes**

H1: A revision-insensitive version of incrementality predicts an immediate truth evaluation on the colour adjective regardless of a potential semantic revision. Increased processing cost is expected for false answers (A,C,D vs. B). In complex contexts (A,D), local answers have to be revised on the preposition in half of the trials (A: 'inside-of' vs. 'outside-of'; D: 'outside-of' vs. 'inside-of').

H2: A revision-sensitive version of incrementality predicts that the position of answer selection differs between simple and complex contexts: For C vs. B, integration difficulties are expected on the adjective, and for A and B, later difficulties are expected when the preposition requires a negative as opposed to an affirmative answer (A: 'outside-of' vs. 'inside-of'; D: 'inside-of' vs. 'outside-of'). These effects should be qualitatively similar to those observed in contexts C vs. B on the adjective.

Predictions regarding the expected ERP components are not entirely straightforward. If context information immediately constrains the upcoming semantic integration, processing difficulties should already be visible in the N400 time-window (Knoeferle et al., 2011; Urbach et al., 2015). The N400 may be followed by a late positive component (Knoeferle et al, 2011). If semantic processing of quantificational restriction is generally delayed, we might expect either absent effects (Kounios & Holcomb, 1992; Urbach & Kutas, 2010), or late positive components reflecting mismatch processing (Kaan et al., 2007, Politzer-Ahles et al., 2013). The detailed predictions are summarized in Table 2.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

TABLE 2. *Predictions*[a]

	Adjective	Preposition
H1	N400 (plus positivity) for A, C, and D vs. B	• Revision: Late positivity for A, *inside-of* vs. *outside-of* Late positivity for D, *outside-of* vs. *inside-of*
H2	N400 (plus positivity) for C vs. B, no effect for A and D	• No effects for B and C • N400 (plus positivity) for A: *outside-of* vs. *inside-of* • N400 (plus positivity) for D: *inside-of* vs. *outside-of*

[a] Although these two hypotheses are the major objectives of Experiment 1, we also conducted additional ad-hoc analyses to obtain a more complete picture of the results obtained. However, despite the statistical significance of such additional outcomes, they must be regarded as post-hoc findings and hence interpreted with great caution because they might simply reflect false-positive results (e.g., Ioannidis, 2005; Ulrich & Miller, 2015). Nevertheless such post-hoc findings may serve as hypotheses for further studies.

As Politzer-Ahles et al. (2013) investigated ERP responses to the quantifier *all*, we also analyzed this position. Due to previously indecisive findings, predictions for this position must be formulated carefully. We either expected a reduced negativity or a sustained positivity when the quantifier did not fit with the preceding picture (i.e., *alle* preceded by contexts (A,B), where a subset reading has to be instantiated).

3.1. METHODS

3.1.1. *Participants*

Twenty-four right-handed students (mean age: 23.5 years) from the University of Tübingen took part in the study (11 male). They were native speakers of German with normal or corrected-to-normal vision and were paid for their participation.

3.1.2. *Materials*

A 4×2-factorial within-subjects design was used with the factors CONTEXT (A–D) and PREPOSITION ('inside-of', 'outside-of'). Pictures were generated as quadruplets via MS PowerPoint. Each picture contained geometrical shapes like triangles and a container form like a circle. Geometrical shapes were both inside and outside of this form. For each quadruplet, the geometrical forms were identical but differed in colour. A total set of 20 quadruplets was generated. Per condition, 20 experimental picture–question pairs were presented, resulting in 160 experimental sentences. To control for strategic effects, we included

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

160 filler questions ending on the adjective. The experiment consisted of 320 sentences. Conditions were evenly spread over 10 blocks. To control for positional effects, two experimental versions were generated. The first block of the first version corresponded to the final block of the second version, the second block of version 1 corresponded to the ninth block of version 2, and so on.

3.1.3. *Procedure*

Participants were seated in a dimly lit, soundproof booth in front of a 17" computer screen. Stimuli appeared in a pseudo-randomized order, in which maximally two items of the same condition appeared in succession. The experimental session was divided into ten blocks (32 trials per block), with breaks between the blocks. At the beginning of each trial, the picture appeared in the centre of the screen for 1500 ms. Then the single words of the sentence were presented word-by-word via rapid serial visual presentation (RSVP) (500 ms per word). The question mark was presented separately in order to avoid predictability of a further restriction. After each question, participants made a truth evaluation ("Did the preceding sentence truly or falsely reflect the content of the picture?"). After the final word had disappeared, three question marks were shown, signalling that participants now had to answer with *wahr* 'true' or *falsch* 'false' by pressing one of two buttons ('F' or 'J') on the keyboard. The keys for true and false answers were counterbalanced across participants. Participants were asked to make their truth evaluation as quickly as possible. The initial above timeout was 1200 ms and was adapted to the response speed of the participants by using an exponentially weighted moving average (Leonhard, Fernández, Ulrich, & Miller, 2011). When participants' reaction times exceeded the current timeout, they received visual feedback (*Schneller!* 'Faster') on the screen. Following the judgement, a blank screen appeared for 500 ms and three exclamation marks in yellow (1200 ms) indicated that participants now could blink until the next picture was presented. A practice session preceded the experimental session. Including electrode application, the experimental session lasted between 2 and 2.5 hours.

3.1.4. *EEG recording*

The EEG was continuously recorded from 32 Ag/AgCl electrodes using a BIOSEMI Active-Two amplifier system: FP1, FP2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO3, PO4, O1, Oz, O2. Six further electrodes were recorded: the electrooculogram (EOG, 4 electrodes) was recorded by means of electrodes placed at the outer canthus of each eye (horizontal EOG) and

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

above and below the participant's left eye (vertical EOG). Two electrodes were put on the left and right mastoid for the purpose of off-line referencing. EEG and EOG recordings were sampled at 1024 Hz during recording, and downsampled to 256 Hz for data analysis. Off-line, electrode sites were referenced to linked mastoids. Raw EEG data were filtered using a 0.3–20 Hz bandpass filter. Before entering the ERP analysis, individual participant data were automatically and manually screened for each trial in order to exclude trials with eye-movement and muscular artefacts. Data per participant and per condition were aggregated from the onset of the critical element (adjective, preposition, and quantifier) to 1000 ms post onset. Afterwards, grand averages were calculated over all participants.[2]

3.1.5. *Data analysis*

For the behavioural data, repeated measures analyses of variance with the within-subject factors CONTEXT (4), and PREPOSITION (2) on mean error rates and reaction times for the truth-value judgements were calculated. Incorrectly answered trials (3.8% of the total trials) were excluded from reaction time and EEG analyses.

ERPs were aggregated from three sentential positions: first, in order to investigate local context effects, brain responses from the onset of the colour adjective were aggregated. Second, potential reanalysis processes were examined from the onset of the preposition. Finally, we aggregated ERPs from the onset of the quantifier. EEG data were analyzed statistically by carrying out repeated-measures ANOVAs for mean amplitude values within time-windows per condition, which were chosen based on visual inspection and previous studies. For the Grand ANOVA, four regions of interest (ROIs) were introduced: left anterior (*Roi 1*: F3, F7, FC1, FC5), right anterior (*Roi 2*: F4, F8, FC2, FC6), left posterior (*Roi 3*: CP1, CP5, C3, P3), and right posterior (*Roi 4*: CP2, CP6, C4, P4).

From the onset of the colour adjective, we carried out ANOVAs with the factors CONTEXT (A,B,C,D) × ROI (*left anterior*, *right anterior*, *left posterior*, *right posterior*). We did not include the Factor PREPOSITION into these analyses, as conditions did not physically differ with respect to this factor on the adjective. From preposition onset, we carried out ANOVAs with the factors CONTEXT (A,B,C,D), PREPOSITION ('inside-of' vs. 'outside-of') × ROI (*left anterior*, *right anterior*, *left posterior*, *right posterior*).

[2] In addition to this method, we also averaged ERPs using the unfiltered raw data and applied a –200 ms pre-stimulus baseline. The qualitative results are highly comparable across both analyses. As we had to manually exclude trials affected by slow drifts, the statistical power of the additional analyses is, however, generally weaker when compared to our original method.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

From quantifier onset, an ANOVA with the factors CONTEXT COMPLEXITY (*simple* (B,C) vs. *complex* (A,D)) was carried out. In order to examine potential strategic effects during the course of the experiment, we carried out additional post-hoc analyses of each of the above-described ANOVAs with the factor EXPERIMENTAL HALF (*first*, *second*).

Individual conditions are only reported if the omnibus ANOVA revealed significant interactions or in case of significant main effects of the factor CONTEXT. The detailed results of all ANOVAs are provided as supplementary materials in the 'Appendix' (available at: <http://dx.doi.org/10.1017/langcog.2016.30>). In the text, we will focus on the most relevant results. Statistical analyses were carried out in a hierarchical manner. Only significant interactions ($p < .05$) were resolved. Corrected p-values (Huynh & Feldt, 1970) were chosen when the analysis involved more than one degree of freedom in the numerator. Probability levels were Bonferroni adjusted for the planned comparisons between the four levels of the factor CONTEXT.

3.2. RESULTS

3.2.1. *Behavioural Data*

Mean error rates and reaction times for the verification task are given in Table 3. Mean error rates in the truth evaluation task were 3.8%, reflecting overall good task performance. Repeated-measures ANOVAs on error rates showed that questions following context D were slightly more erroneous than all other context–sentence pairs. A main effect of CONTEXT was found ($F(3,69) = 34.17$; $p < .001$). Whereas contexts B and D did not differ from each other ($p = .3$), all other contexts did (all p values $> .03$). Reaction times showed an interaction between CONTEXT and PREPOSITION ($F(3,69) = 2.8$; $p = .05$), which was due to the fact that the two prepositions differed for context D ($F(1,23) = 4.34$; $p < .05$; all other p values $< .2$). We refrain from interpreting reaction times in detail as they were not directly timelocked to the disambiguating positions.

3.2.2. *Colour adjective*

Figure 2 shows ERPs from the onset of the colour adjective. Conditions differed within an early (300–400 ms) and a late (450–800 ms) time-window. The detailed results of the ANOVA are provided as supplementary materials in the 'Appendix' (A1.1).

Statistical analyses in the *early time window* revealed a significant main effect of CONTEXT ($F(3,69) = 22.2$; $p < .001$), as well as an interaction between ROI and CONTEXT ($F(9, 207) = 5.74$; $p < .05$). Whereas complex contexts (A,D) did not differ from each other ($p = 1$), all other contexts

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

TABLE 3. *Behavioural results for the single conditions in Experiment* 1

	Verification results	
Condition	Correct responses (%)	Reaction times (ms)
A *inside-of*	96.1	294.2
A *outside-of*	98.6	273.4
B *inside-of*	99.0	271.7
B *outside-of*	91.3	284.1
C *inside-of*	96.5	254.5
C *outside-of*	98.4	281.3
D *inside-of*	99.2	275.6
D *outside-of*	90.1	256.3

did, with significant differences in ROI 3 and ROI 4 (all p values < .05). Finally, no interactions with the factor HALF were observed (all F values < 2.07). In the *late time window,* an effect of CONTEXT (p < .001) was found. Whereas complex conditions did not differ (p = 1), all other conditions did (all p values < .01). A ROI and CONTEXT interaction (p = .001) reveals variances in effect sizes in the single ROIs. No interactions with the factor HALF were observed (all F values < 2.16).

In sum, the statistical analyses confirmed that false answers already differed from true answers on the colour adjective when contextual information allowed for an unambiguous decision: for simple contexts, false answers (C) elicited a negativity followed by a late positivity when compared to true answers (B). The complex contexts (A,D) did not differ from each other, but differed from both of the simple contexts.

3.2.3. *Preposition*

Figure 3 shows ERPs from preposition onset for contexts A to D. Comparably to the adjective, conditions differed within an early (300–400 ms) and a late (450–800 ms) time-window. These differences were restricted to previously ambiguous conditions following complex contexts (A,D). The complete results are provided as supplementary materials in the 'Appendix' (A1.2). Statistical analyses in the *early time window* showed an effect of CONTEXT (F(3,69) = 4.32; p < .01), as well as a three-way interaction between ROI, CONTEXT, and PREPOSITION (F(9,207) = 7.52; p < .001): whereas an effect of PREPOSITION was found for contexts A and D in each ROI (all p values < .01; with the exception of context A in the left posterior ROI with p = .07), no significant effects were observed for previously unambiguous conditions following simple contexts (B,C) (all p values > .83). The post-hoc analyses revealed an interaction between ROI, CONTEXT, and HALF

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

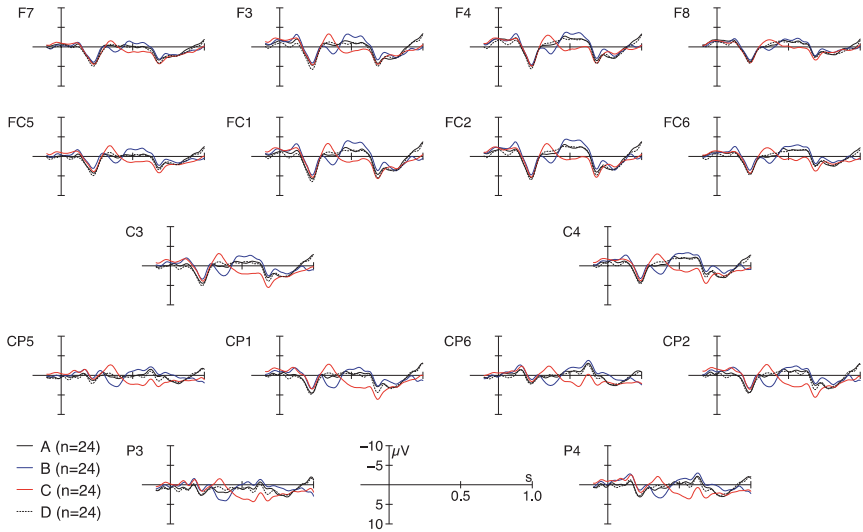


Fig. 2. Grand average ERPs from the onset of colour adjective, Experiment 1.

($F(9,207) = 3.65$; $p < .01$), which became significant in the left-posterior ROI, ($F(3,69) = 2.91$; $p < .05$; all other ROIs: $F < 2.19$). However, the effect of CONTEXT was not significant in any of the experimental halves (first half: ($F(3,69) = 2.41$; $p = .30$; second half: ($F(3,69) = 2.34$; $p = .32$). In the *late time window*, an effect of CONTEXT was found ($F(3,69) = 15.04$; $p < .001$), as well as a three-way interaction between ROI, CONTEXT, and PREPOSITION ($F(9,207) = 3.25$; $p < .01$). Significant effects were restricted to the *left posterior* ROI. No interactions with the factor HALF were observed (all F values < 2.51).

In sum, the statistical analyses from the onset of the preposition revealed significant effects only for complex contexts (A,D). For those cases, a similar biphasic ERP pattern for false answers was found as the one on the colour adjective, i.e., an early negativity that was followed by a late positivity.

3.2.4. *Quantifier*

Figure 4 shows ERPs from the onset of the quantifier. Based on Politzer-Ahles et al. (2013), we analyzed three time-windows: an early (250–500 ms), a late (500–1000), and a long time-window (300–1000 ms). No significant effects were found in the early time-window (all p values > .2). In the late time-window, an effect of CONTEXT COMPLEXITY was observed ($F(1,23) = 7.44$; $p < .05$). In the long time-window, no significant effects were found (all p values > .09). In the additional analysis, no effects were found (all F values < 1.96). As a whole,

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

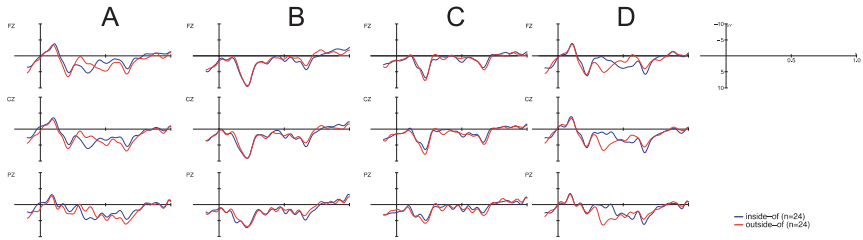


Fig. 3. Grand average ERPs from the onset of the preposition for the midline electrodes, Experiment 1.

the analysis of the quantifier position revealed that, in a late time-window, complex conditions were slightly more positive than simple conditions.

3.3. DISCUSSION

The present experiment used a picture–question verification task to test whether context information incrementally constrains the processing of questions involving quantifier restriction. We examined whether a semantic commitment was made on the earliest possible position in the sentence (H1) or whether parsing decisions were delayed in the risk of a semantic revision (H2). As earlier studies had either observed an N400 for context–sentence mismatches or a late positivity (or a combination of both), we were also interested in the specific processing phase affected by the incongruence.

The present results are compatible with H2. On the colour adjective, a biphasic ERP pattern was observed for false as opposed to true answers (following context C vs. B). In line with previous studies, a mismatch between pictorial and semantic information was reflected by a negativity that was followed by a late positivity (Knoeferle et al., 2011). Under H1, the complex contexts (A,D) were not supposed to differ from context C: in each of these conditions, an immediate local commitment would have resulted in a 'false' answer. Thus, a similar negativity would have been expected on the adjective for A, C, and D. By contrast, the amplitude of the complex contexts differed from the simple contexts. Whether this pattern signals a locally under-specified semantic representation can only be decided by considering the results at the preposition, where an unambiguous disambiguation becomes available.

On the preposition, effects were restricted to the complex conditions. For complex conditions, the incongruent conditions (A: 'outside-of' vs. 'inside-of'; D: 'inside-of' vs. 'outside-of') elicited a comparable biphasic pattern, as did the incongruent conditions for the simple conditions on the colour adjective. This finding is again in accord with H2. Thus, the truth evaluation in the

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

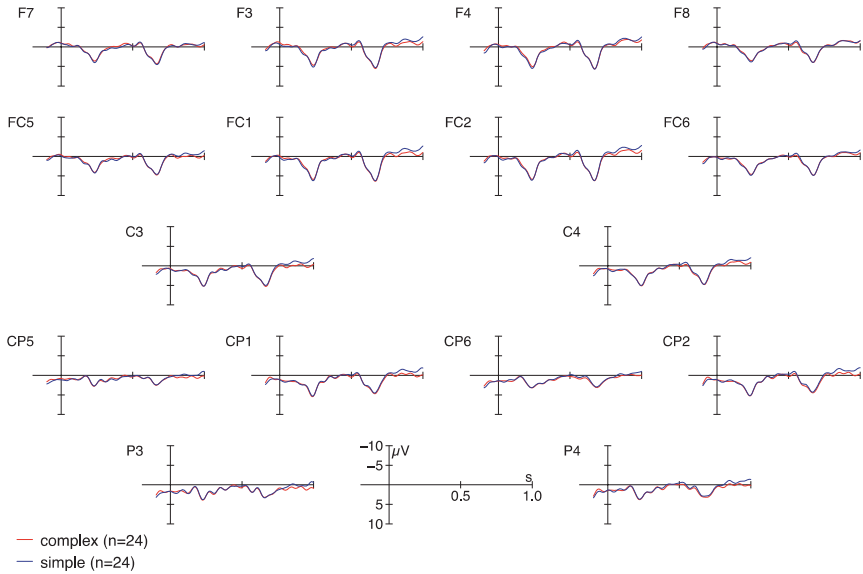


Fig. 4. Grand average ERPs from the onset of the quantifier, Experiment 1.

complex conditions was delayed until the position of the preposition, with a semantically under-specified representation before the ultimate point of disambiguation. Under H1, the directly opposite pattern would have been expected for complex contexts: if a semantic evaluation had already been made on the colour adjective, increased processing costs would have occurred for a shift in local truth evaluation (A: 'inside-of' vs. 'outside-of'; D: 'outside-of' vs. 'inside-of'). The positivity following the negativity is descriptively smaller than the one on the adjective and is statistically significant only in one Ro1. One explanation for this finding could be that the specific information that has to be falsified differs between simple and complex contexts. For simple contexts, falsification involves the properties of all pictorial objects, i.e., the total set of blue or red objects. By contrast, for complex sentences, only a subset of the available objects is affected by the falsification process.

Finally, our additional analyses showed differences between quantifiers following simple and complex contexts. In contrast to Politzer-Ahles et al. (2013), we found a late positivity for the complex conditions. Perhaps most plausibly, this effect reflects processing costs associated with the instantiation of a subset reading, as it becomes clear at this position that the sentence could only be correct if a further restriction will follow. The late effect is principally in accord with results from Kaan et al. (2007). However, as no neutral baseline (such as *all* or *they*) was presented in that study, these results do not provide decisive evidence for or against processing costs associated with a subset reading.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

In sum, the present results show that contextual effects may incrementally constrain the processing of questions involving quantifier restriction. At the earliest fully disambiguating position (the colour adjective for simple contexts, and at the preposition for complex contexts), false answers elicited a negativity followed by a positivity. Thus, a truth evaluation already occurs at an early processing stage and is made as soon as contextual information is unambiguously available. Considering previous studies on picture–sentence verification, both the negativity and the late positivity had sometimes been associated with strategic effects rather than reflecting 'real' semantic processing (Knoeferle et al., 2011; Vissers et al., 2008). According to that view, the negativity reflects attentional mismatch detection rather than a fully-fledged linguistic-semantic analysis (D'Arcy & Connolly, 1999). Moreover, a long-standing debate in the literature has been concerned about whether late positivities in language tasks reflect linguistic processes (e.g., Friederici, Hahne, & Saddy, 2002; Osterhout & Holcomb, 1992), or are members of the P300 family, reflecting domain-general mechanisms (Coulson, King, & Kutas, 1998). According to the latter view, late positivities reflect increased attentional demands when the current stimulus is particularly relevant for fulfilling the specific task requirements. For examining whether attentional processes affected the present results, we conducted a second experiment employing a probe detection task, in which attention was directed away from the explicit mapping between the context picture and the target question.

4. Experiment 2

The second study intended to replicate the findings from Experiment 1 using a probe detection task that should prevent readers from focusing on the mapping between the context picture and the target question. In addition, we provided a post-Experiment questionnaire, in which participants indicated whether they had read the questions as a whole or whether they had focused on specific sentential positions. If so, they were asked to indicate the positions they focused on.

4.1. METHODS

4.1.1. *Participants*

Twenty-four right-handed students (mean age: 26.3 years) from the University of Tübingen took part in the study (12 male). They were German native speakers with normal or corrected-to-normal vision and were paid for their participation. Due to ocular or motor artefacts, two participants were excluded from the final analyses.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

4.1.2. *Materials*

The identical context–question pairs as in Experiment 1 were presented. Filler items were mainly identical to those in Experiment 1. In half of the fillers, we used pictures in which two different geometrical figures appeared. The total number of experimental trials was 320. For the probe detection, five probe positions were randomly assigned: the picture, the first noun, the colour adjective, the preposition, and the final noun. Probes were either identical (50%) or non-identical to the previously presented picture or word. For non-identical pictures, probes either differed with respect to the geometrical objects or to the colours. Linguistic probes were semantically related to the words in the probe position.

4.1.3. *Procedure*

The experimental settings and the presentation of trials were identical to Experiment 1 up to the question mark. 300 ms after question mark offset, a visual probe (a picture or a word) was presented, and participants had to decide whether the probe had already been shown in the current trial. Participants answered with 'yes' or 'no' (50%) by pressing one of two answer buttons ('F' or 'J'). The keys for the 'yes' and 'no' answers were counterbalanced across participants. For timeouts, we applied the same procedure as in Experiment 1. In 15.6% of the trials, a box appeared instead of the probe (*Bitte beantworte die Frage* 'Please answer the question'), and participants had to answer the question by pressing the 'yes' or 'no' button. No reaction time feedback was provided for these trials. Following the decision or answer, a blank screen appeared (500 ms), and three exclamation marks in yellow (1200 ms) were shown, indicating that participants could blink until the picture was presented. A practice session preceded the experimental session. A post-Experiment questionnaire was filled in after the ERP study.

4.1.4. *EEG recording and data analysis*

EEG recording and data analysis were identical to Experiment 1.

4.2. RESULTS

4.2.1. *Behavioural data*

Mean error rates and reaction times for the probe detection task are given in Table 4. Mean error rates in the probe detection task were 21.5%. Repeated-measures ANOVAs on error rates showed an interaction between CONTEXT and PREPOSITION ($F(3,63) = 11.26$; $p < .001$).

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

TABLE 4. *Behavioural results for the single conditions in Experiment* 2

	Probe detection results	
Condition	Correct responses (%)	Reaction times (ms)
A *inside-of*	80.3	838.0
A *outside-of*	72.7	881.5
B *inside-of*	78.2	802.4
B *outside-of*	83.3	819.4
C *inside-of*	81.9	831.6
C *outside-of*	74.3	868.2
D *inside-of*	75.9	871.4
D *outside-of*	81.2	805.8

For contexts A and C, the preposition *innerhalb* 'inside-of' showed more correct responses than *außerhalb* 'outside-of' (A: $F(1,21) = 10.03$; $p < .01$; C: $F(1,21) = 8.98$; $p < .01$). For contexts B and D, the preposition *innerhalb* caused more errors (B: $F(1,21) = 6.33$; $p < .05$; D: $F(1,21) = 8.76$; $p < .01$). Reaction times showed an effect of CONTEXT ($F(3,63) = 5.46$; $p < .01$), and an interaction between CONTEXT and PREPOSITION ($F(3,63) = 6.45$; $p = .001$). Prepositional differences were restricted to complex contexts (A: $F(1,21) = 4.41$; $p < .05$; D: $F(1,21) = 10.86$; $p < .01$; B and D: each F value < 0.4). As in Experiment 1, we will refrain from interpreting reaction times. Nevertheless, they do appear to mirror the pattern of ERP results reported below.

4.2.2. *Post-Experiment questionnaire*

From the 22 participants who entered the final analysis, 18 (81.8%) reported to have read the whole question. The remaining four participants reported to have focused their attention on the probe positions.

4.2.3. *ERP data*

Analogous to Experiment 1, we analyzed ERPs from the onset of the colour adjective, the preposition, and the quantifier. Given the high error rates in the present study and the fact that the probe detection was not related to the picture–sentence mapping, we did not exclude errors from the ERP analysis.

4.2.4. *Colour adjective*

Figure 5 shows ERPs from the onset of the colour adjective. Conditions again differed within an early (300–400 ms) time-window. In contrast to

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

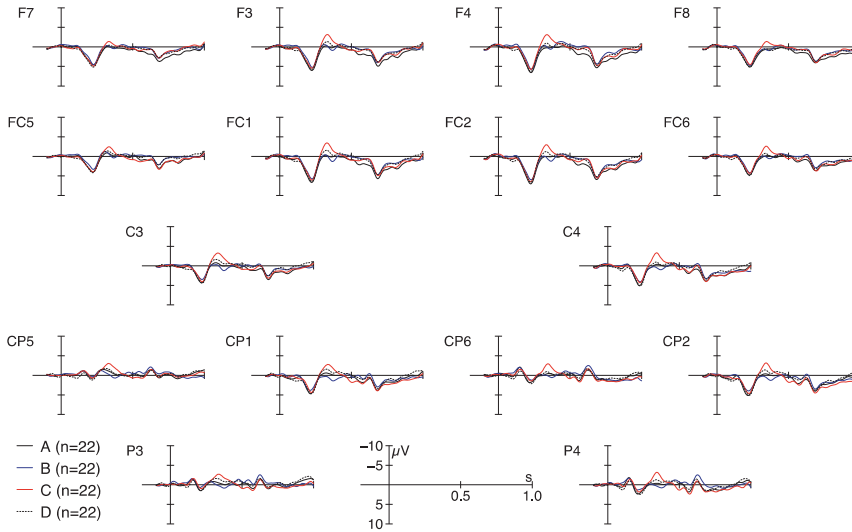


Fig. 5. Grand average ERPs from the onset of colour adjective, Experiment 2.

Experiment 1, no differences were found in the late time-window (450–800 ms). The detailed results are provided as supplementary materials in the 'Appendix' (A2.1).

The statistical analyses in the *early time-window* (300–400 ms) revealed an effect of CONTEXT ($F(2,46) = 10.68$; $p < .001$), as well as a ROI × CONTEXT interaction ($F(6,138) = 2.63$; $p < .05$). In contrast to the first study, condition A did not differ from condition B (all F values = 1). No interactions with the factor HALF were observed (all F values < 1.88). In the *late time-window* (450–800 ms), no significant differences or interactions were found (all F values < 2.5). In sum, statistical analyses on the adjective confirmed that the experimental contexts elicited differential brain responses on the position of the colour adjective. Comparably to Experiment 1, false answers yielded the most negative effect, and true answers were most positive. Though the complex conditions did not differ significantly from each other, condition A slightly differed from condition B.

4.2.5. *Preposition*

Figure 6 shows ERPs from the onset of the preposition. Again, conditions differed within an early (300–400 ms) time-window. No differences were descriptively visible in the late (450–800 ms) time-window. The detailed statistical results are provided as supplementary materials in the 'Appendix' (A2.2).

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

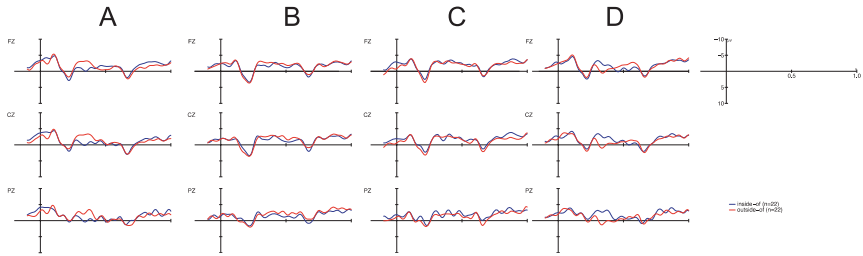


Fig. 6. Grand average ERPs from the onset of the preposition for the midline electrodes, Experiment 2.

Comparably to Experiment 1, the statistical analyses in the *early time-window* (300–400 ms) showed that the difference between the preposition types affected the previously ambiguous complex contexts. An effect of CONTEXT was found ($F(3,69) = 6.63$; $p = .001$), as well as a three-way interaction between ROI, CONTEXT, and PREPOSITION ($F(9,207) = 3.23$; $p < .01$). Whereas an effect of PREPOSITION was found for contexts A and D in each ROI (all p values < .01), no significant effects were observed for contexts B and C (all p values > .11). No effects or interactions with the factor HALF were observed (all F values < 1.59). In contrast to Experiment 1, no effects of PREPOSITION or PREPOSITION × CONTEXT interactions were found in the *late time-window* (450–800 ms). However, a main effect of CONTEXT was found ($F(3,63) = 11.65$; $p < .05$), as well as a CONTEXT × ROI interaction ($F(9,189) = 2.34$; $p < .05$). No effects or interactions with the factor HALF were observed (all F values < 2.11).

4.2.6. *Quantifier*

Figure 7 shows ERPs from the onset of the quantifier. Again, we analyzed an early (250–500 ms), a late (500–1000 ms), and a long (300–1000 ms) time-window.

No significant effects were found in any of the time-windows (all ps > .11). The additional analyses showed an interaction with HALF (late time-window: $F(1,21) = 4.51$; $p < .05$; long time-window: $F(1,21) = 4.58$; $p < .05$). Whereas no context effects were observed in the first half of the experiment (all F values < 1.83), contexts differed in the second half of the experiment (all F values > 2.60).

4.3. DISCUSSION

The present study was intended to replicate the findings from Experiment 1 by using a task that distracted attention away from the picture–question match.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

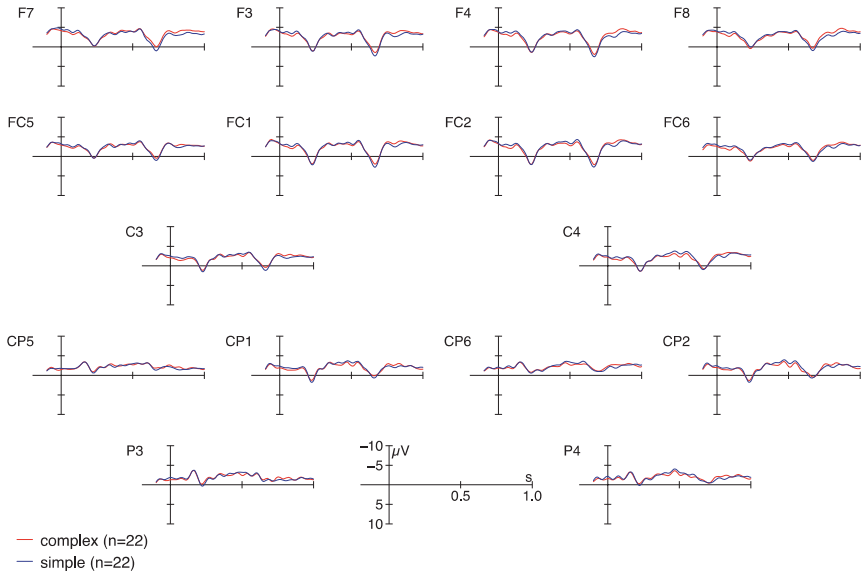


Fig. 7. Grand average ERPs from the onset of the quantifier, Experiment 2.

With respect to ERP effects in the early time-window, the present findings replicated the results from Experiment 1. An early negativity was observed on each of the disambiguating positions. On the adjective, the negativity was found for false answers (C) as opposed to true answers (B). Again, the complex conditions (A,D) differed from these simple conditions. The early negativity on the preposition was again restricted to the previously ambiguous conditions. Thus, even without a task-relevant match between context and target, a comparable pattern as in Experiment 1 was observed: in the cases of a safe decision, the truth-value of the question was computed as quickly as possible. By contrast, in the cases of a risky decision, the parser evaluated the truth-value only at the earliest unambiguous position, i.e., at the preposition. The finding of an early negativity is in accord with Hypothesis H2 and indicates that compositional-semantic processing may rapidly affect sentence comprehension at early processing phases, thus ruling out a purely strategy-based alternative explanation of the findings of Experiment 1.

In contrast to Experiment 1, no consistent effects were found in the late time-window. On the colour adjective, the single conditions did not differ from each other. Thus, the experimental task manipulation affected the late positivity, which was modulated by attentional demands in Experiment 1, but not in Experiment 2. On the preposition, conditions differed in the late time-window. Crucially, these effects were not related to the picture–question mapping but showed up as contextual effects, which plausibly can

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

be attributed to the fact that questions following simple contexts already differed from those following complex contexts before the preposition was processed. Visual inspection of the ERPs indicates that the complex conditions are indeed already more positive than the simple conditions from the onset of the preposition and could thus be spillover effects from the adjectival region (Steinhauer & Drury, 2012). In the current experiment, effects on the quantifier position were restricted to the second half of the experiment, thus indicating that processing difficulties associated with a subset interpretation may have been modulated by attentional demands.

Finally, results from our post-Experiment questionnaire revealed that the vast majority of participants read the questions in their entirety.[3] Importantly, none of the participants who were not reading the whole sentences reported to have been exclusively paying attention to the colour adjective or the preposition.

5. General discussion

The present study investigated pictorial effects on the compositional-semantic processing of questions involving quantificational restriction. We tested whether the meaning of the universal quantifier is incrementally restricted by contextual information regardless of the risk of a potentially upcoming restriction. Moreover, we examined whether this process is modulated by attentional demands. In our first experiment, a picture–question verification task, attention towards the mapping between the context picture and the target question was directly relevant for task fulfilment. Our second experiment employed a probe detection task, in which this mapping was not relevant for successfully fulfilling the task.

Most generally, our results show that the restrictive properties of quantifier meaning constrain question processing at early processing phases. Across two tasks, a negativity (300–400 ms) was found when the pictorial information was not in accord with a positive answer. The negativity was followed by a late positivity (450–800 ms) in Experiment 1, where the task required an explicit mapping between the pictorial context and the question. In Experiment 2, a probe detection task, attention was directed away from this match, and no late positivity was observed. In both experiments, the negativity was elicited only when a commitment of the parser did not result in the risk of a semantic revision. Thus, context effects were directly evaluated on the colour adjective when preceded by unambiguous pictorial contexts. However, when a potentially following restriction could lead to a shift in

[3] Visual inspection did not reveal any differences with regard to the observed ERP response between groups.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

semantic meaning, a comparable negativity was observed only from the onset of the preposition. As a whole, the observed pattern of results is compatible with a revision-sensitive strategy (H2), in which the parser only makes a semantic commitment in safe contexts.

Our results are in line with previous studies investigating context effects on semantic processing (Hagoort & van Berkum, 2007). A deteriorated semantic fit was reflected by a negative ERP component, which was followed by a positivity. Contrary to previous studies (Nieuwland, 2016; Urbach et al., 2015), the negativity could not have been triggered by lexical expectancies. First, after having processed the context picture, any kind of colour adjective could principally follow in the question. Second, the experimental design had been balanced in such a way that each context occurred with both possible colours. For instance, context B was followed by the adjective 'blue', but it was also followed by the adjective 'red' in another trial, for which it then constituted context C. Thus, in addition to studies on lexical integration, the present study indicates that also higher-level compositional-semantic processes have early effects on sentence comprehension (Hunt et al., 2013; Knoeferle et al., 2011; Vissers et al., 2008). One novelty of the current experiments is the finding that an incremental semantic commitment depends on the reliability of contextual information, which may affect the risk of revision due to alternative possibilities of domain restriction.

In contrast to the negativity, the occurrence of the late positivity was restricted to Experiment 1, where the experimental task encouraged an attentional focus towards the picture–question match. Interestingly, a comparable effect was also discussed in Nieuwland (2016; see also Freunberger & Nieuwland, 2016), who compared quantificational interpretation in a verification task and a task that was not directly related to the target sentences. Comparably to our results, the positivity exclusively occurred in the verification task. Therefore, the current positivity seems to be related to task demands rather than reflecting semantic processing per se.

Though examining the quantifier region was not the primary intention of the present study, the late positivity for the subset reading in Experiment 1 is partly in accord with previous findings (Kaan et al., 2007; Politzer-Ahles et al., 2013). In the remainder of the 'General discussion', we will discuss the observed ERP pattern in detail.

5.1. THE EARLY NEGATIVITY

Given the early onset of the negativity, the question arises whether the observed effect is in fact a 'semantic' N400 or whether it reflects lower-level processes related to stimulus properties. Early negativities could alternatively

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

be viewed to be N2b components reflecting an attentional sensitivity towards the match between visual and linguistic information rather than a fully-fledged semantic analysis (Knoeferle et al., 2011; Vissers et al., 2008). Standardly, the N2b occurs in oddball tasks, in which voluntary attention is directed towards mismatch detection (Pritchard, Shappell, & Brandt, 1991), and has been suggested to reflect a deviation from an active template elicited by preceding stimuli (Näätänen, 1995). In the present experiments, such an attentional mechanism would involve a succession of different processing steps, such as analyzing the visual input, keeping this information active in working memory, and matching each incoming word with the current template (D'Arcy & Connolly, 1999). When encountering contradictory information, a representational mismatch would elicit the N2b, and would therefore reflect an unsuccessful mapping between the lexical properties of the input word and the pictorial representation.

Considering the simple contexts, such a strategy could straightforwardly explain the observed results. For instance, after having processed the context picture B, the pictorial representation contains information about the colour ('blue') and the presented objects ('triangles'), and probably the fact there are blue triangles inside and outside of a circle. This representation needs to be upheld in memory until matching or mismatching information will be encountered, and is then taken out of the memory buffer. When arriving at the adjective, the colour 'blue' matches part of the representation, whereas the colour 'red' leads to a mismatch between the pictorial representation and the target word. Regarding the complex contexts, the pictorial information contains information about both colours ('blue' and 'red'), the presented objects, and information about the subsets (the red or the blue one). Under a purely attention-driven account, it is somewhat difficult to explain why, on the colour adjective, complex contexts do not elicit comparable mismatch effects as simple contexts. One explanation for this finding could be that the actually encountered colour information is part of the current pictorial representation and refers to one of the subsets involved, hence, no mismatch is perceived. However, under such an account, it is not entirely clear why any effects were observed at the preposition at all. At this position, the encountered preposition types are also included in the current representation. Nevertheless, clear-cut mismatch effects were observed for the complex contexts. Moreover, the current negativity reliably occurred across different tasks. Whereas the picture–question verification task explicitly focused participants' attention towards the mapping between the picture and the sentence, fulfilment of the probe detection task was principally possible without instantiating any relation between the picture and the question. In addition, the post-Experiment questionnaire revealed that the majority of participants had read the whole questions without focusing on specific words, thus indicating that the observed

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

ERP pattern may relate to processes involved in question processing instead of reflecting a pure attentionally driven match.

An alternative explanation could thus be that it is indeed an N400 reflecting higher-level semantic processing. The early onset of the N400 could have been evoked by a particularly high semantic expectancy elicited by the context picture (Vissers et al., 2008). According to this view, a presented picture elicits a strong semantic expectancy that a following sentence will correctly describe the pictorial content. This expectancy is violated when an incoming word is in conflict with a correct interpretation. However, note that, in the present experiment, questions instead of declarative sentences were presented. As a consequence, there is no principled reason to assume that comprehenders expect a 'yes' rather than a 'no' answer. In semantic theory, questions have sometimes been taken to denote sets of possible answers (Hamblin, 1973). Thus, the *yes/no* question *Are all triangles blue?* is taken to denote the set {all triangles are blue, it is not the case that all triangles are blue}. Under this approach, the questions in the present experiments should denote exactly the same set of propositions independently of the utterance context. Given that there is currently little ERP evidence on the on-line comprehension of questions, we can only speculate why 'no' answers elicited an N400. One explanation would be that negative answers are generally more complex, as they involve an additional negation step (Kaup, Lüdtke, & Zwaan, 2006). Negative answers would thus involve an instantiation of a negative tag into the actually processed meaning representation (Carpenter & Just, 1975), or the activation of two different states of affairs (both the negated and an actual state). For instance, when processing *the door is not closed*, both a closed and an open door could be assumed to be simulated (Lüdtke, Friedrich, De Filippis, & Kaup, 2008). Adopted to the present scenario, such an account would predict that, for instance, when processing 'blue' after context C, both the originally shown context C as well as the correct alternative would need to be simulated, whereas following context B, only the blue context picture would be simulated.

In ERP studies on negation processing it is as yet an unresolved issue whether negative sentences are processed in an immediate or delayed fashion. While studies employing verification paradigms report absent (Fischler et al., 1983) or delayed effects (Lüdtke et al., 2008; Wiswede, Koranyi, Müller, Langer, & Rothermund, 2013), other studies report immediate N400 effects indistinguishable from non-negated controls (Nieuwland & Kuperberg, 2008). In accord with the latter findings, the brain response towards negative answers in the present experiments already occurred at an early time-window. This is expected under Nieuwland and Kuperberg's pragmatic account of negation processing because, for our polar questions, both answer alternatives were pragmatically equally licensed.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

An alternative explanation for the preference for 'yes' answers could be that the semantic processing of questions is constrained by predictive processes. In that case, contextually given cues elicit the prediction that a speaker asking a question will refer to the already presented information instead of one of the numerous possible alternatives. For instance, a picture including only blue objects would elicit the prediction that the speaker will ask about the properties already specified instead of introducing a new reference set (e.g., a set of red objects). Such a proposal is generally compatible with previous work on lexical predictability in declarative clauses (Brothers, Swaab, & Traxler, 2015; Nieuwland, 2016; Roehm, Bornkessel-Schlesewsky, Roesler, & Schlesewsky, 2007; Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2010). For instance, Roehm et al. (2007) manipulated lexical predictability and semantic relatedness in an antonym study, and observed that both of these factors have an impact on processes within the N400 time-window. Whereas semantic relatedness affected the N400, lexical predictability (e.g., of the word *white* in a sentential context like *The opposite of black is* …) elicited an early positive deflection in the same time-window. Co-occurring effects in the N400 time-window thus may reflect the simultaneous processing of qualitatively distinct processes related to predictive processes and the semantic evaluation of the currently processed word (see Lau, Holcomb, & Kuperberg, 2013, for discussion). Adopted to the present findings, a processing advantage of unambiguous 'yes' answers would signal that correct answers had already been predicted prior to arriving at the critical region. Thus, specific conceptual features of a following restrictor (e.g., a specific colour of the adjective) would already be added to the current representation of each newly incoming word, and would then be evaluated as soon as a compatible or incompatible input is encountered. If such an approach is on the right track, it still needs to be determined why questions do not differ from declarative clauses with respect to their expected restrictors. In particular, in the Roehm et al. (2007) example, a specific lexical item (i.e., *white*) is predicted, and any other item would render the whole statement false. By contrast, questions do not assert statements and do not have truth-values that could be falsified by unexpected items. Therefore, an interesting task for follow-up studies would be to further investigate differences between declaratives and questions manipulating different linguistic and non-linguistic context types. Thus, while the underlying mechanism for the preference of 'yes' answers in questions still needs to be determined, both of the above-mentioned accounts are compatible with the finding of ERP differences in the N400 time-window (see also Hunt et al., 2013, for comparably early effects in a picture–sentence verification paradigm on the processing of the quantifier *some*).

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

In sum, the present negativity is partly compatible with both an N2b approach, or could alternatively reflect an N400 component. According to the former view, the present pattern of results reflects an attentionally driven representational match between the context picture and the target question, whereas an N400 approach indicates that the negativity reflects increased processing cost associated with the generation of a negative answer.

5.2. THE LATE POSITIVITY

As already mentioned in the discussion of Experiment 1, a long-standing debate in the literature concerns the question whether late positivities in linguistic tasks uniquely reflect linguistic processing (Friederici et al., 2002; Osterhout & Holcomb, 1992) or whether they need to be related to the P300 family (Coulson et al., 1998). Recently, it has been proposed that late positivities involved in well-formedness evaluations can also be captured by referring to the P300 (Bornkessel et al., 2011). According to this view, late positivities in linguistic tasks are related to binary decision processes. For instance, classifying linguistic stimuli into well-formed and ill-formed structures requires such a binary process. According to so-called 'P600-as-P3' approaches, stimuli that are highly relevant for fulfilling the requirements of a given task often elicit late positivities that reflect increased attentional demands (Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014). Instead of being language-related, the P300 is thus supposed to reflect domain-general processing strategies.

Our results are compatible with such an approach. The verification task in Experiment 1 involved a binary decision process, and the truth evaluation could have increased the attention towards the picture–question mapping. The finding that the P300 already occurred at the mismatch positions indicates that a truth evaluation was made as soon as the anomaly became available. In the first experiment, the task-relevant decision could already have been made during the processing of the critical sentential positions (i.e., on the adjective for contexts B and C, and on the preposition for contexts A and D), as, at those positions, all relevant information for task fulfilment (e.g., the pictorial representation as well as the linguistic information) was already available. Short reaction times for the clause-final verification task additionally indicate that the truth evaluation had been made prior to the clause-final response. By contrast, in Experiment 2, the actual probe detection could only have been made after the final word of the sentence had been processed. Up to the presented probe, participants could not guess which part of the sentence would be relevant for task fulfilment. As a consequence, increased processing costs in terms of generally longer reaction times were found.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

As a whole, the present positivity varied as a function of the attention towards the picture–question match across experiments and is thus compatible with current P600-as-P3 approaches.

6. Conclusion

The present study shows that pictorial context information may constrain early processing stages of compositional-semantic processing. In addition to previous studies, the current results also indicate that an incremental semantic meaning evaluation is dependent on the reliability of contextual information. In the present experiments, a semantic commitment was only made when contextual information was unambiguous and thus did not lead to a risk of a potential semantic revision. Finally, our results indicate that the experimental task used in studies on semantic processing can affect the overall pattern of ERP results. In line with previous studies, a mismatch between contextual and semantic information triggered an early negativity, which was not affected by differences in attentional demands across the two tasks. By contrast, a late positivity accompanying this negativity was only observed when the task required increased attention towards the mapping between contextual and semantic information.

Supplementary materials

For supplementary material for this paper, please visit <http://dx.doi.org/10.1017/langcog.2016.30>.

REFERENCES

Augurzky, P. (2006). Attaching relative clauses in German – the role of implicit and explicit prosody in sentence processing. *MPI Series in Human Cognitive and Brain Sciences*, 77, Leipzig.

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics & Philosophy*, **4**(2), 159–219.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011). Think globally: cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, **117**(3), 133–152.

Bott, O., & Schlotterbeck, F. (2015). The processing domain of scope interaction. *Journal of Semantics*, **32**(1), 39–92.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: prediction takes precedence. *Cognition*, **136**, 135–149.

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

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: a psycholinguistic model of verification. *Psychological Review*, **82**, 45–76.

Coulson, S., King, J. W., & Kutas, M. (1998). ERPs and domain specificity: beating a straw horse. *Language and Cognitive Processes*, **13**(6), 653–672.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

D'Arcy, R. C. N., & Connolly, J. F. (1999). An event-related brain potential study of receptive speech comprehension using a modified Token Test. *Neuropsychologia*, **37**, 1477–1489.
Delong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, **8**, 1117–1121.
Deschamps, I., Agmon, G., Loewenstein, Y., & Grodzinsky, Y. (2015). The processing of polar quantifiers, and numerosity perception. *Cognition*, **143**, 115–128.
Dotlačil, J., & Brasoveanu, A. (2015). The manner and time course of updating quantifier scope representations in discourse. *Language, Cognition and Neuroscience*, **30**(3), 305–323.
Dwivedi, V., Phillips, N. Einagel, S., & Baum, S. (2010). The neural underpinnings of semantic ambiguity and anaphora. *Brain Research*, **1311**, 93–109.
Eckstein, K., & Friederici, A. D. (2006). It's early: event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience*, **18**, 1696–1711.
Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: evidence from the N400. *Psychophysiology*, **45**(4), 554–558.
Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. J. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, **20**, 400–409.
Fodor, J. D. (2002). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *Proceedings of the North East Linguistic Society*, 32 (pp. 113–132). Amherst, MA: GSLA, University of Massachusetts.
Frazier, L., Clifton, Jr., C., Rayner, K., Deevy, P., Koh, S., & Bader, M. (2005). Interface problems: Structural constraints on interpretation? *Journal of Psycholinguistic Research*, **34**, 193–223.
Freunberger, D., & Nieuwland, M. S. (2016). Incremental comprehension of spoken quantifier sentences: evidence from brain potentials. *Brain Research*, **1646**, 475–481.
Friederici, A. D., Hahne, A., & Saddy, D. (2002): Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, **31**, 45–63.
Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: speech acts* (pp. 41–58). New York: Academic Press.
Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, **304**, 438–441.
Hagoort, P., & van Berkum, J. J. A. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*, **362**, 801–811.
Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, **10**, 41–53.
Hartshorne, J., Snedeker, J., Liem Azar, S., & Kim, A. (2015). The neural computation of scalar implicature. *Language, Cognition and Neuroscience*, **30**(5), 620–634.
Huang, Y., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: insight into the semantics–pragmatics interface. *Cognitive Psychology*, **58**, 376–415.
Hunt, L., Politzer-Ahles, S., Gibson, L., Minai, U., & Fiorentino, R. (2013). Pragmatic inferences modulate N400 during sentence comprehension: evidence from picture–sentence verification. *Neuroscience Letters*, **534**, 246–251.
Huynh, H., & Feldt, L. S. (1970). Conditions under which the mean square ratios in repeated measurement designs have exact F-distributions. *Journal of the American Statistical Association*, **65**, 1582–1589.
Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, **2**, e124.
Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: evidence from brain potentials. *Journal of Memory and Language*, **86**, 157–171.
Kaan, E., Dallas, A. C., Barkley, C. M. (2007). Dealing with new discourse referents: an ERP study. *Brain Research*, **1146**, 199–209.
Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, **38**, 1033–1050.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

Knoeferle, P., Urbach, T. P., & Kutas, M. (2011). Comprehending visual context influences on incremental sentence comprehension: insights from ERPs and picture–sentence verification. *Psychophysiology*, **48**, 495–506.

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, **121**, 459–479.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, **62**, 621–647.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 161–163.

Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics*, 2nd ed. (pp. 659–724). New York: Elsevier.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. K. (2013). Dissociating N400 effects of prediction from association in single word contexts. *Journal of Cognitive Neuroscience*, **25**, 484–502.

Leonhard, T., Ruiz Fernández, S., Ulrich, R., & Miller, J. (2011). Dual-task processing when Task 1 is hard and Task 2 is easy: Reversed central processing order? *Journal of Experimental Psychology: Human Perception and Performance*, **37**, 115–136.

Lüdtke, J., Friedrich, C., De Filippis, M., & Kaup, B. (2008). ERP correlates of negation. *Journal of Cognitive Neuroscience*, **20**, 1355–1370.

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear Hear*, **16**(1), 6–18.

Nieuwland, M. S. (2016). Quantification, prediction, and the online impact of sentences truth-value: evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **42**, 316–334.

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: an ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, **63**(3), 324–346.

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: an event-related potential study on the pragmatics of negation. *Psychological Science*, **19**, 1213–1218.

Nouwen, R. (2010). What's in a quantifier? In M. Everaert, T. Lentz, H. de Mulder, O. Nilsen, & A. Zondervan (Eds.), *The linguistics enterprise: linguistik aktuell 150* (pp. 235–256). Amsterdam: John Benjamins.

Noveck, I. A., & Sperber, D. (2007). The why and how of experimental pragmatics: the case of 'scalar inferences'. In N. Burton-Roberts (Ed.), *Advances in pragmatics*. Basingstoke: Palgrave.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, **31**, 785–804.

Otten, M., Nieuwland, M. S., & van Berkum, J. J. A. (2007). Great expectations: specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, **8**, 89.

Paterson, K. B., Filik, R., & Moxey, L. M. (2009). Quantifiers and discourse processing. *Language and Linguistics Compass*, **3**, 1390–1402.

Peters, S., & Westerståhl, D. (2006). *Quantifiers in language and logic*. Oxford: Oxford University Press.

Politzer-Ahles, S., Fiorentino, R., Jiang, X., & Zhou, X. (2013). Distinct neural correlates for pragmatic and semantic meaning processing: an event-related potential investigation of scalar implicature processing using picture–sentence verification. *Brain Research*, **1490**, 134–152.

Pritchard, W. S., Shappell, S. A., & Brandt, M. E. (1991). Psychophysiology of N200/N400: a review and classification scheme. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (pp. 43–106). Greenwich, CT: JAI press.

Roehm, D., Bornkessel-Schlesewsky, I., Roesler, F., & Schlesewsky, M. (2007). To predict or not to predict: influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, **19**, 1259–1274.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, **137**, 29–39.
Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgment task: evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, **31**(6), 817–840.
Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, **120**(2), 135–162.
Ulrich, R., & Miller, J. (2015). p-hacking by post hoc selection with multiple opportunities: detectability by skweness test? Comment on Simonsohn, Nelson, and Simmons (2014). *Journal of Experimental Psychology: General*, **144**, 1137–1145.
Urbach, T. P., DeLong, K. A., & Kutas, M. (2015). Quantifiers are incrementally interpreted in context, more than less. *Journal of Memory and Language*, **83**, 79–96.
Urbach, T. P., & Kutas, M. (2010). Quantifiers more or less quantify online: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, **63**, 158–179.
van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **31**, 443–467.
Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, **22**, 1682–1700.
Villalta, Elisabeth. (2003). The role of context in the resolution of quantifier scope ambiguities. *Journal of Semantics*, **20**, 115–162.
Vissers, C. T., Kolk, H. H., van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: evidence from ERPs in a picture–sentence matching task. *Neuropsychologia*, **46**(4), 967–982.
Westerståhl, D. (1985). Determiners and context sets. In J. van Benthem & A. ter Meulen (Eds.), *Generalized quantifiers in natural language* (pp. 45–71). Dordrecht: Foris.
Wijnen, F., and Kaan, E. (2006). Dynamics of semantic processing: the interpretation of bare quantifiers. *Language and Cognitive Processes*, **21**(6), 684–720.
Wiswede, D., Koranyi, N., Müller, F., Langer, O., & Rothermund, K. (2013). Validating the truth of propositions: behavioral and ERP indicators of truth evaluation processes. *Social Cognitive and Affective Neuroscience*, **8**(6), 647–653.

https://doi.org/10.1017/langcog.2016.30 Published online by Cambridge University Press