Neuroscience Letters 630 (2016) 141–146

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Research article

These lemons are sour: Investigating the influence of demonstrative determiners on the N400 complex

Carolin Dudschig*, Claudia Maienborn, Barbara Kaup

University of Tübingen, Germany

HIGHLIGHTS

- Investigation of specifying-determiner integration into the N400 complex.
- Offline plausibility shows clear influence of determiners on comprehension.
- The N400 complex does not reflect these offline meaning differences.
- No instant influence of determiner-related plausibility on the N400 complex.

ARTICLE INFO

Article history:
Received 14 May 2016
Received in revised form 11 July 2016
Accepted 18 July 2016
Available online 25 July 2016

Keywords:
Language
N400
Determiners
Lexical association
LSA
Predictability

ABSTRACT

Demonstrative determiners create a joint focus of attention of speaker and hearer on a particular set of entities in a given context. Despite their omnipresence, very little is known with regard to the online integration of demonstrative determiners during comprehension. Here we investigated whether the N400 reflects meaning differences introduced by determiners. We analyzed whether previously established differences – between correct, contingent, world-knowledge and semantic violations in bare plural generic sentences – were modified by the addition of demonstrative determiners. Off-line ratings suggested that the correct sentences in the generic version (e.g. *Zebras are stripy*) become less sensible when adding a demonstrative determiner (e.g. *These zebras are stripy*). In contrast, contingent sentences – that is, sentences with predicates that are true for a subgroup of exemplars but not for all – become more sensible than correct sentences if a demonstrative determiner is added (e.g. *These trousers are stripy*). Both semantic and world-knowledge violations slightly improve in sensibility if demonstrative determiners are added (*These ladybirds/journeys are stripy*). None of these differences was reflected in the N400 amplitude. Instead, the previous findings using the bare plural generic sentences were fully replicated, showing largest N400 amplitudes for semantic violations, followed by contingent and world-knowledge violations and finally the correct sentences. This suggests that demonstrative determiners are not automatically integrated during meaning comprehension. Implications with respect to the processes that are reflected in the N400 will be discussed.

1. Introduction

Definite determiners – such as *the* or the demonstrative *this* or *these* – are omnipresent, however very little is understood about how these are integrated during language comprehension. Determiners are functional elements that combine with a noun phrase and specify its referential properties. Whereas a bare plural as in *Lemons are sour* makes a generic statement about the whole class of lemons, a demonstrative as in *These lemons are sour* directs the hearer's attention towards a particular subset of lemons present in a given context (see Ref. [5]). In the current study, we compared the influence of the demonstrative determiner *these* during online language comprehension to its offline impact on sentence-plausibility ratings. Despite demonstrative determiners having a core communicative function, they only rarely have been the focus of language comprehension research.

Many studies recently investigated the influence of quantifying determiners (e.g. *some, most, all*) on online language comprehension, and focused on the question whether their influence on offline meaning comprehension is reflected in early online measures [27]. Early studies [19] showed that the N400 is not sensitive to meaning differences triggered by quantifier manipulation (e.g.

* Corresponding author at: Schleichstr. 4, 72076 Tübingen, Germany.
E-mail address: carolin.dudschig@uni-tuebingen.de (C. Dudschig).

http://dx.doi.org/10.1016/j.neulet.2016.07.031
0304-3940/

Some/No/All dogs are animals). This suggests that rather than an online integration of the quantifying determiner taking place, a semantic category priming (dog → animal) is underlying the N400 effect. Similar results – that is a lack of influence on the N400 complex – are typically reported for negation integration (e.g., [10,11]). However, in pragmatically licensed negative sentences (e.g., *In moderation, drinking red wine isn't good/bad* . . .) negation integration seems to take place instantly [28,30]. A study by [36] showed that also quantifiers can be – at least partially – integrated into online comprehension. For sentences such as *Most/few farmers grow worms* versus *Most/few farmers grow crops* the offline sentence ratings predicted a full crossover interaction, the 'few-farmers-worms' condition should be as easy to processes as the 'most-farmers-crops' condition. The results showed that the N400 was indeed affected by the correct quantifier but did not result in a full crossover interaction. Thus, the authors concluded that some – but not all – aspects of quantifier interpretation are conducted online. However, there are recent studies suggesting that under specific conditions quantifier interpretation can take place fully incrementally. Urbach et al. [35] showed that supporting discourse contexts (e.g., *Alex was an unusual toddler. Most/Few kids prefer sweets/vegetables*. . .) can lead to a full cross-over interaction of typical/untypical noun and the quantifier condition in the N400 under specific task and reading condition. Similarly, violations of pragmatic informativeness introduced by the use of quantifiers (e.g., *Some people have lungs/pets*) can influence the N400 but only given participants' high pragmatic abilities and contextual factors that put focus on the local under informativeness [29] (for similar results in a study using counterfactuals see: [20]). Nieuwland [27] showed that under conditions of high predictability (e.g., *Most/Few gardeners plant their flowers during the spring/winter*) quantifying determiners are also fully integrated during online comprehension.

In the current study we will also use the N400 measure to investigate the impact of demonstrative determiners on online and offline sentence comprehension. Demonstrative determiners have a core communicative function reflected by the fact that demonstratives are among the first and most frequent expressions of child language (e.g. [3]) and the fact that they are truly universal (e.g. [6]). Additionally, demonstrative determiners are more frequent in language use than quantifiers. Also from a linguistic perspective one can argue that demonstrative determiners are less complex than quantifying determiners. On a Fregean approach, demonstrative determiners pattern with definite determiners in yielding individual denoting noun phrases (whose denotation is of type e; see e.g., [8,33]). Quantifying determiners such as *some, all, most*, in contrast, yield noun phrases of a more complex logical type («e,t>,t). Thus demonstrative determiners can be viewed as less complex than quantifying determiners. In virtue of this difference and the cognitive differences outlined above, demonstrative determiners might be differently integrated during comprehension than quantifying determiners. Therefore, it is worth looking at them as a different phenomenon – with the potential of an instant online integration even in non-predictable reading conditions without context. There is one study investigating definite and demonstrative determiner integration into the N400 complex Prasada et al. [41]. In their study the authors found a main effect of the determiner (e.g., *Diamonds are expensive/cheap* vs. *This diamond is expensive/cheap*), showing that the N400 generally decreased in sentences with a demonstrative in contrast to generic sentences. However, the authors did not find an interaction between the version of the sentence (generic vs. specific) and the adjective (characteristic vs. uncharacteristic) leaving open whether the use of plurals in the generic sentences triggered the N400 effect.

There continues to be debate about the processes that contribute to the N400 component [22] (e.g., [24]). On the one side, several studies suggest that lexical factors influencing word retrieval from memory play a central role. These factors include non-semantic factors such as word frequency, but also associations between single words and even between larger contexts and single words [21,38]. Also, lexical associations as measured via latent semantic analysis (LSA; [23]) have been shown to influence the N400 amplitude [37]. On the other side, there are studies suggesting that the N400 reflects sentential integration processes (e.g., [2]). In a strict sense, such an integration view assumes that the N400 itself reflects combinatorial processes taking place at the sentence level, an assumption still highly debated. Probably the most accepted view of the N400 at the moment is driven by the evidence that the N400 amplitude is influenced by the predictability of the incoming words and whether an incoming word shares – for example semantic or phonological features – with the predicted or preactivated word (see Refs. [17,27,25,26]). In this view all information that is available to the comprehender is used to predict – and subsequently integrate – upcoming words. Hereby, both the sentence's constraints and the comprehender's skills influence the integration process reflected in the N400 amplitude (e.g., [29]). Despite the ongoing debate regarding the exact mechanisms underlying the N400, the N400 is a standard measure to investigate mechanisms at work during language comprehension.

In particular, the N400 was employed in studies investigating whether comprehension takes place in a fully incremental manner (all information is integrated at once) – or whether comprehension is partly incremental, whereby a first and a second step of meaning calculation takes place. A famous example has been the study published in *Science* by Hagoort et al. [15] investigating whether linguistic knowledge is integrated earlier than non-linguistic world-knowledge during language comprehension. Consider the following example: Both the sentence *Journeys are stripy* and the sentence *Ladybirds are stripy* are evaluated as being wrong. However, the reasons why these sentences are considered wrong are different from a linguistic perspective. The predicate *stripy* demands an expression that denotes a concrete physical object with a surface. It can therefore not be combined with an eventive expression such as *journey*. For the sentence *Ladybirds are stripy*, this purely linguistic demand is not violated. However, we know from our experiences that ladybirds are dotted and not stripy. In Hagoort et al.'s study both violation types resulted in similar N400 onset latencies which challenged the view that linguistic and non-linguistic knowledge sources are integrated at different time points during comprehension. In the current study, we will make use of the N400 amplitude findings previously established by Hagoort et al. [15] and replicated by Dudschig et al. [7]. We will use the identical sentences as in [7], however, we will now implement demonstrative determiners instead of bare plurals. Whereas bare plurals yield generic statements, adding a demonstrative determiner leads to a specific statement. In the original study, correct sentences (*Zebras are stripy*), contingent sentences where the predicate is true for some but not all exemplars (*Trousers are stripy*), world-knowledge violation sentences (*Ladybirds are stripy*) and semantic violation sentences (*Journeys are stripy*) were used. The semantic violations resulted in largest N400 amplitudes followed by world-knowledge violations and contingent sentences, and smallest N400 amplitudes were elicited by correct sentences. Indeed, these sentences nicely allow the investigation of determiner integration. In offline measures, the contingent sentences with the addition of the determiner (*These trousers are stripy*) should result in best plausibility measures, followed by originally correct sentences (*These zebras are stripy*) and world-knowledge violations (*These ladybirds are stripy*), whereas semantic violations (*These journeys are stripy*) should not be rated as particularly plausible. We expect correct sentences to become less plausible than contingent sentences because they are underinformative – see Ref. [12] maxime of informativeness – and thereby violate pragmatic

Table 1
Mean sensibility rating from the offline rating separately for the specific and the generic versions of the sentences.

Violation-Type	Correct (SD)	Contingent (SD)	Semantic (SD)	World-Know (SD)
Specific	5.53 (1.45)	6.31 (0.79)	1.61 (0.62)	2.67 (1.14)
Generic	6.42 (0.41)	4.19 (0.96)	1.36 (0.47)	1.67 (0.85)

Note: Mean ratings and the standard deviations (SD).

expectancies: Given the fact that *all* zebras are stripy, why state that *these* specific zebras are stripy? (see Refs.: [29,19]). If online N400 measures reflect early integration of determiners – resulting in detection of underinformativeness – then we would expect the original N400 result [7] to be reversed in the current study. Specifically, we would expect that the contingent sentences result in smaller N400 amplitudes than the correct sentences, mirroring the results from the expected offline measures. In contrast, if the N400 rather reflects basic word-level effects and is not influenced by underinformativeness [19] we expect identical results to the study by [7] and to see no influence of the determiner.

2. Method

2.1. Participants

Forty native speakers of German participated for course credit or payment. Four participants were excluded due to too few trials remaining after artifact correction (<70%). Thirty-six participants were included in the analysis (M_{age} = 21.44, SD = 5.39, 8 male, all right-handed).

2.2. Material

The materials were adopted from Dudschig et al. [7]. In that study 160 critical sentences and 180 plausible filler sentences were constructed in German (+20 practice sentences). The critical sentences were divided into four subgroups. 40 sentences were correct sentences – these sentences originally reflected true generic statements but now a demonstrative determiner was added (e.g. *These zebras are stripy*). 40 sentences were constructed that involved a predicate that was true for some members of the category but not for all – the contingent condition (e.g. *These trousers are stripy*). 40 sentences contained semantic violations (e.g. *These journeys are stripy*). Finally, another set of 40 sentences contained a world-knowledge violation (e.g. *These ladybirds are stripy*). All experimental sentences consisted of a noun phrase, the copula *sein* ('to be') and an adjective. Various filler sentences had a different sentence structure. The nouns in the experimental sentences did not differ with regard to log-frequency across conditions (Leipziger Wortschatzportal), $F(3, 156) = 2.09$, $p = 0.10$. The critical words were the adjectives (identical across all conditions). The semantic association between the nouns and the adjectives was measured using LSA with the LSA-fun R-package [13,14] using the SdeWaC web-based corpus with ~800 million words [9]. Not all words were in the matrix, reducing the ANOVA's degrees of freedom. The main effect of condition was significant, $F(3, 123) = 10.96$, $p < 0.001$ ($M_{correct} = 0.46$, $M_{contingent} = 0.44$, $M_{world\text{-}knowledge} = 0.43$, $M_{semantic} = 0.26$). Post-hoc Welch's t-tests of adjoined conditions showed a significant difference only between the world-knowledge and the semantic condition, $t(55.58) = 4.55$, $p < 0.001$. In order to check for predictability, a cloze test was performed. Twenty-two participants (M_{age} = 23.52, SD = 2.30, 7 males) had to generate a final adjective given the sentence beginning. The ANOVA showed a main effect of condition, $F(3, 156) = 32.94$, $p < 0.001$ ($M_{correct} = 34\%$, $M_{contingent} = 7\%$, $M_{world\text{-}knowledge} = 0\%$, $M_{semantic} = 0\%$). Post-hoc Welch's t-tests showed a difference between the correct and the contingent condition, $t(56.16) = 4.75$, $p < 0.001$, and between the world-knowledge and the contingent condition, $t(39.17) = 3.00$, $p < 0.01$. Overall, the cloze test showed that the correct condition was particularly highly predicted, and the LSA measure showed that the semantic condition was particularly low associated.

2.3. Offline Rating

The critical sentences were rated by 80 participants (M_{age} = 25.83, SD_{age} = 7.01, 22 male) in a web-based study according to sensibility from 1 ('does not make sense') to 7 ('does fully make sense'). Half of the participants read the sentences in the generic version (no determiner), the other half in the specific version (with determiner). An ANOVA including the factors violation-type (correct, contingent, semantic and world-knowledge) and sentence-type (generic vs. specific) showed a significant main effect of violation-type, $F(3, 234) = 631.37$, $p < 0.001$ (see Table 1 for means). Additionally, there was a main effect of sentence-type, $F(1,78) = 22.74$, $p < 0.001$, suggesting that overall generic sentences were rated as less sensible ($M = 3.41$) than specific sentences ($M = 4.03$). Most importantly, there was a significant interaction between violation-type and sentence-type, $F(3, 234) = 51.29$, $p < 0.001$. These differences were analyzed in more detail using Welch's t-tests. In the generic version, the correct sentences were rated as most sensible, followed by the contingent sentences (correct vs. contingent: $t(52.87) = 13.50$, $p < 0.001$), the word-knowledge violations (contingent vs. world-knowledge: $t(76.90) = 12.38$, $p < 0.001$) and finally the sentences in the semantic-violation condition (world-knowledge vs. semantic: $t(60.44) = 2.07$, $p < 0.05$). For the specific sentences, in line with our hypothesis, the contingent sentences were rated as most sensible, followed by the correct sentences (contingent vs. correct: $t(60.31) = -3.00$, $p < 0.01$), the world-knowledge violation (correct vs. world-knowledge: $t(74.04) = 9.80$, $p < 0.001$) and finally the sentences in the semantic-violation condition (world-knowledge vs. semantic: $t(60.06) = 5.13$, $p < 0.001$). Most importantly, the correct and the contingent sentences swapped their position in the sensibility rating between the generic and the specific version. This is also reflected in an interaction between violation-type (correct vs. contingent) and sentence-type, $F(1,78) = 97.11$, $p < 0.001$.

2.4. Procedure and design

The procedure and design was identical to the original study [7]. Each trial started with a black fixation-cross presented in the center of the screen for 1500 ms. The screen background was set to a light grey color. Each word was displayed centrally in black for 300 ms. Words were separated by a 300 ms blank screen. The last word was followed by a 1000 ms blank screen. Comprehension questions were asked in approximately 8% of the trials. Participants had to indicate with a button press on a standard keyboard (n or v key) whether a statement (e.g., *All crocodiles are able to swim*) is correct or not according to the preceding sentence (e.g., *Crocodiles can go into the water*). The experiment was split into one practice block and four experimental blocks. The experimental blocks always contained 40 critical sentences and 45 filler sentences. The 40 critical sentences consisted of ten sentences of each condition.

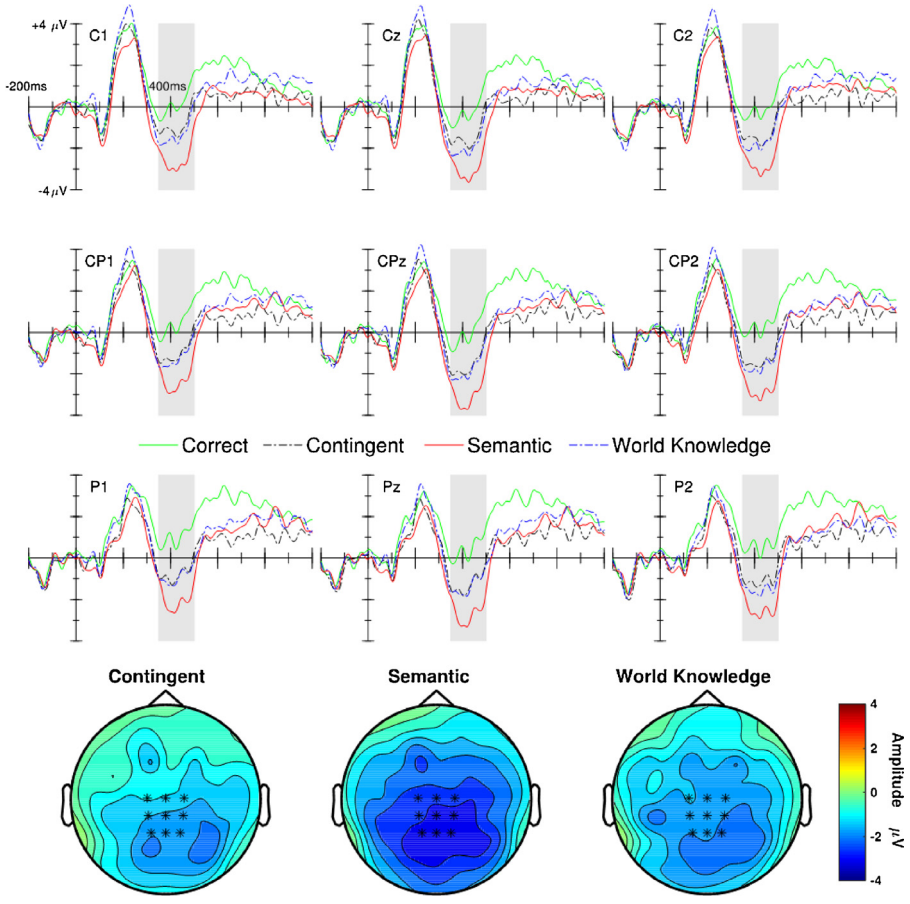

Fig. 1. Mean activity recorded over centroparietal electrodes (C1, Cz, C2, CP1, CPz, CP2, P1, Pz, P2).

2.5. EEG recording and analysis

Electrophysiological measures. EEG activity was recorded at 256 Hz continuously from midline electrodes: Fpz, AFz, Fz, FCz, Cz, CPz, Pz, POz, Oz, and Iz; left hemisphere electrodes: IO1, Fp1, AF3, AF7, F1, F3, F5, F7, F9, FC1, FC3, FC5, FT7, C1, C3, C5, M1, T7, CP1, CP3, CP5, TP7, P1, P3, P5, P7, PO3, PO7, O1 and homologous right hemisphere electrode sites. ERP analysis was performed using MATLAB toolboxes (EEGLAB: [4]; FieldTrip: [31]) and custom MATLAB scripts. Off-line, all EEG channels were re-referenced to the average of the mastoid electrodes and high-pass filtered (0.1 Hz, 30 dB/oct). Next, artifacts were removed or corrected in a procedure similar to the one implemented by Nolan et al. [40] and described in previous studies (e.g., Dudschig et al. [39]). First, epochs containing extreme values in single electrodes (values larger ±1000 μV) were removed, as were trials containing values exceeding ±75 μV in multiple electrodes (if unrelated to EOG activity). Secondly, z-scored variance measures were calculated for all electrodes, and noisy EEG electrodes (z-score > ±3) were removed if their activity was uncorrelated to EOG activity. Thirdly, this cleaned data was subjected to an ICA-analysis [1]. ICA components representing ocular activity were identified using z-scored measures of the absolute correlation between the ICA component and the hEOG and vEOG activity, subsequently confirmed by visual inspection [18]. Fourthly, previously removed noisy channels were interpolated in the ICA-cleaned data set using the average activity of adjacent channels (distance 4 cm, ~3–4 neighbors). Finally, a low-pass filtered (30 Hz, 36 dB/oct) was applied.

2.6. Statistical analysis

The mean N400 ERP amplitudes from the centroparietal electrodes (C1, Cz, C2, CP1, CPz, CP2, P1, Pz, P2) were determined in the time window from 350 ms to 500 ms separately for each condition and aligned to a 100 ms pre-stimulus baseline. A repeated measure ANOVA with the factor violation-type (correct, contingent, semantic vs. world-knowledge) was conducted in order to analyze the N400 and a sphericity correction was applied.

3. Results

The critical word elicited a positive P2 component followed by a negative deflection peaking at around 400 ms, the N400 component (see Fig. 1). An ANOVA of the mean amplitude in the N400 time window showed a main effect of violation-type, $F(3, 105) = 17.40$, $p < 0.001$ ($M_{correct} = 0.28$ μV, $M_{contingent} = -1.22$ μV, $M_{semantic} = -2.65$ μV, $M_{world} = -1.48$ μV). Follow-up single comparisons were calculated. There was a significant difference between the correct and the contingent sentences, $F(1, 35) = 11.37$, $p < 0.01$, and between correct and the world-knowledge violation sentences, $F(1, 35) = 15.81$, $p < 0.001$. The contingent and world-knowledge violation did not differ from each other, $F < 1$. Finally, both the contingent and the world-knowledge condition differed significantly from the semantic violation, $F(1, 35) = 15.69$, $p < 0.001$ and $F(1, 35) = 9.80$, $p < 0.01$, respectively. These findings replicate the finding from earlier studies [7]. We also conducted a between experiment comparison, comparing the results from the previous study using the bare plural generic sentences with the results of the

current study. This analysis showed no interaction between experiment and violation-type ($F < 1$), confirming the conclusion that the determiner did not modify the N400 amplitude.

4. Discussion

Definite and demonstrative determiners are typically used for referring to a particular subset of entities given in the context. Despite their crucial communicative function only little is understood with regard to their online integration during language comprehension. Here we aimed at closing this gap. As expected, the offline sensibility ratings showed that contingent bare plural sentences benefit most – with regard to their sensibility – from adding a demonstrative determiner (*These trousers are stripy*). Most importantly, the contingent sentences become even more sensible than their correct counterparts (*These zebras are stripy*). These offline sentence-sensibility ratings predicted that a full interpretation of the determiners should result in the full crossover of the contingent and the correct condition in the N400 complex. However, this offline meaning difference is not reflected in the N400 online measure. In contrast, the N400 amplitude results fully mirrored previous results investigating contingent, semantically violated and world-knowledge violated variants of bare plural generic sentences. These results suggest that demonstrative determiners – despite being linguistically and cognitively less complex than quantifying determiners – still show no online influence on the N400 complex.

The cloze measurement indicated that the sentence endings of the correct condition were still more predictable than the endings of the more plausible contingent sentences. This suggests, that for contingent sentences – whose endings are not very predictable – the N400 does not seem to reflect meaning differences introduced by definite determiners (that are clearly reflected in the offline plausibility ratings). This interpretation is in line with previous studies investigating the integration of the more complex quantifying determiners and negation into the N400 complex that show that plausibility typically does not influence the N400 independent of predictability (e.g., [27,19]. Also, if context pragmatically licenses the target sentences then quantifying determiners are integrated more easily [35]. Therefore, future studies would be needed to investigate the positive influence of context and predictability with regard to the incremental integration of demonstrative determiners. Critically, there was also a significant difference in predictability between the contingent and the world-knowledge violation condition, which was not reflected in any N400 effect. Therefore, a simple predictability hypothesis can also not explain the data pattern, especially as it is known that even very small differences in predictability have the potential to affect comprehension [34].

Another related issue with the current experimental setup refers to the pragmatics of demonstrative determiners. A determiner such as *these* leads the reader's attention to a particular set of entities that is given by the context of the utterance. However, in the current experimental setup the sentences were presented out of the blue without providing further context information. Therefore, the participants were not able to evaluate the truth of the sentences. Thus, one might argue that the sentences were pragmatically infelicitous. However, the sentence ratings showed a main effect of sentence-type: Overall the sentences with the addition of the demonstrative determiners were rated as more plausible than their bare plural generic counterparts. This suggests, that there was no major problem with the sentences due to the demonstrative determiners. In any case, even if the lack of context was a problem for the processing of demonstrative determiners, then one might still argue that this problem existed consistently across all four conditions and therefore should not impact the main findings. Another related issue is the use of a rather high amount of violated sentences. In the current study we used an untypical amount of correct filler sentences in order to reduce the true/violated ratio. Nevertheless, it is possible that under experimental situations participants stop attempting to process subtle meaning differences necessary to support full interpretation of the demonstrative determiners. In contrast, in the plausibility rating participants are forced to attend to sentence meaning more precisely than when just reading sentences and answering comprehension questions.

The current finding might also be relevant to the discussion regarding one- vs. two-step models of sentence comprehension. Specifically, the current study suggests that the N400 amplitude difference between the world-knowledge and the semantic violation conditions persists even in situations where the sentence structure does not prompt checking the sentence's truth value. Whereas bare plural generic sentences may be truth-evaluated without additional context information (*Ladybirds are stripy*), adding a demonstrative determiner (*These ladybirds are stripy*) should reduce the tendency in the readers to actively evaluate the truth of the sentences (but see Refs. [16,32]). Another aspect with regard to the amplitude difference between the semantic and the world-knowledge condition is the lexical association between the noun and the adjective in these conditions. The LSA analysis using rather large corpora shows a difference between the semantic and the world-knowledge condition, which suggests that N400 amplitude differences between these conditions might be explained by such word associations. However, there was no difference in the LSA between the correct and the contingent condition, suggesting that these differences must be due to other sources. Maybe, if sentence endings are predictable in some way, then predictability indeed reduces the N400 size (correct/contingent condition), whereas if sentence endings are rather untypical, other types of processes determine the N400 size. However, such an explanation still leaves open why there was no difference in the N400 between the contingent and the world-knowledge condition – given the differences in predictability. Taken together, these explanations are highly post-hoc and future studies would be needed to investigate the influence and interplay of different – probably not always independent – measures (e.g., LSA, predictability, context, etc.) on the N400 complex with the ultimate goal to understand the exact processes underlying the N400.

Acknowledgements

This project was supported by the Institutional Strategy of the University of Tübingen (German Research Foundation, ZUK 63) and a RISC grantappointed to Carolin Dudschig (Ministry of Science, Research and the Arts Baden-Württemberg), by the SFB833/B4 project of Barbara Kaup and Hartmut Leuthold, and by the SFB833/A1 project of Claudia Maienborn (German Research Foundation). We thank two anonymous reviewers for very helpful comments on this manuscript.

References

[1] A.J. Bell, T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (6) (1995) 1129–1159.
[2] C. Brown, P. Hagoort, The processing nature of the N400: evidence from masked priming, J. Cognit. Neurosci. 5 (1) (1993) 34–44.
[3] E.V. Clark, From gesture to word: on the natural history of deixis in language acquisition, in: J.S. Bruner, A. Garton (Eds.), Human Growth and Development, Oxford University Press, Oxford, 1978, pp. 85–120.
[4] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods 134 (1) (2004) 9–21.

[5] H. Diessel, Deixis and demonstratives, in: Claudia Maienborn, Klaus von Heusinger, Paul Portner (Eds.), Semantics. An International Handbook of Natural Language Meaning, de Gruyter, Berlin, 2012, pp. 2407–2432.
[6] R.M.W. Dixon, Demonstratives: a cross-linguistic typology, Stud. Lang. 27 (2003) 61–122.
[7] C. Dudschig, C. Maienborn, B. Kaup, Is there a difference between stripy journeys and stripy ladybirds? The N400 response to semantic and world-knowledge violations during sentence processing, Brain Cogn. 103 (2016) 38–49.
[8] P. Elbourne, Demonstratives as individual concepts, Ling. Philos. 31 (4) (2008) 409–466.
[9] G. Faaß, K. Eckart, SdeWaC – a corpus of parsable sentences from the web, in: Language Processing and Knowledge in the Web, Springer, Berlin Heidelberg, 2013, pp. 61–68.
[10] H.J. Ferguson, A.J. Sanford, H. Leuthold, Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds, Brain Res. 1236 (2008) 113–125.
[11] I. Fischler, P.A. Bloom, D.G. Childers, S.E. Roucos, N.W. Perry, Brain potentials related to stages of sentence verification, Psychophysiology 20 (4) (1983) 400–409.
[12] P. Grice, Logic and conversation, in: P. Cole, J. Morgan (Eds.), Syntax and Semantics, 3: Speech Acts, Academic Press, New York, NY, 1975, pp. 41–58.
[13] F. Günther, C. Dudschig, B. Kaup, LSAfun-An R package for computations based on latent semantic analysis, Behav. Res. Methods 47 (2015) 930–944.
[14] F. Günther, C. Dudschig, B. Kaup, Latent semantic analysis cosines as a cognitive similarity measure: evidence from priming studies, Q. J. Exp. Psychol. 69 (2016) 626–653.
[15] P. Hagoort, L. Hald, M. Bastiaansen, K.M. Petersson, Integration of word meaning and world knowledge in language comprehension, Science 304 (5669) (2004) 438–441.
[16] M.B. Isberner, T. Richter, Does validation during language comprehension depend on an evaluative mindset? Discourse Process. 51 (2014) 7–25.
[17] A. Ito, M. Corley, M.J. Pickering, A.E. Martin, M.S. Nieuwland, Predicting form and meaning: evidence from brain potentials, J. Mem. Lang. 86 (2016) 157–171.
[18] T.P. Jung, S. Makeig, C. Humphries, T.W. Lee, M.J. Mckeown, V. Iragui, T.J. Sejnowski, Removing electroencephalographic artifacts by blind source separation, Psychophysiology 37 (02) (2000) 163–178.
[19] J. Kounios, P.J. Holcomb, Structure and process in semantic memory: evidence from event-related brain potentials and reaction times, J. Exp. Psychol.: Gen. 121 (4) (1992) 459–479.
[20] E. Kulakova, M.S. Nieuwland, Pragmatic skills predict online counterfactual comprehension: evidence from the N400, Cognit. Affective Behav. Neurosci. (2016) 1–11.
[21] M. Kutas, K.D. Federmeier, Electrophysiology reveals semantic memory use in language comprehension, Trends Cognit. Sci. 4 (12) (2000) 463–470.
[22] M. Kutas, S.A. Hillyard, Reading senseless sentences: brain potentials reflect semantic incongruity, Science 207 (4427) (1980) 203–205.
[23] T.K. Landauer, P.W. Foltz, D. Laham, An introduction to latent semantic analysis, Discourse Process. 25 (2-3) (1998) 259–284.
[24] E.F. Lau, C. Phillips, D. Poeppel, A cortical network for semantics:(de) constructing the N400, Nat. Rev. Neurosci. 9 (12) (2008) 920–933.
[25] S. Laszlo, K.D. Federmeier, A beautiful day in the neighborhood: an event-related potential study of lexical relationships and prediction in context, J. Mem. Lang. 61 (3) (2009) 326–338.
[26] M.S. Nieuwland, The truth before and after: brain potentials reveal automatic activation of event knowledge during sentence comprehension, J. Cognit. Neurosci. 27 (2015) 2215–2228.
[27] M. Nieuwland, Quantification, prediction, and the online impact of sentences truth-value: evidence from event-related potentials, J. Exp. Psychol.: Learn. Mem. Cognit. 42 (2016) 316–334.
[28] M.S. Nieuwland, G.R. Kuperberg, When the truth is not too hard to handle an event-related potential study on the pragmatics of negation, Psychol. Sci. 19 (2008) 1213–1218.
[29] M.S. Nieuwland, T. Ditman, G.R. Kuperberg, On the incrementality of pragmatic processing: an ERP investigation of informativeness and pragmatic abilities, J. Mem. and Lang. 63 (2010) 324–346.
[30] M.S. Nieuwland, A.E. Martin, If the real world were irrelevant: so to speak: the role of propositional truth-value in counterfactual sentence comprehension, Cognition 122 (2012) 102–109.
[31] R. Oostenveld, P. Fries, E. Maris, J.M. Schoffelen, FieldTrip: open source software for advanced analysis of MEG EEG, and invasive electrophysiological data, Comput. Intell. Neurosci. (2011) 156869.
[32] T. Richter, S. Schroeder, B. Wöhrmann, You don't have to believe everything you read: background knowledge permits fast and efficient validation of information, J. Pers. Soc. Psychol. 96 (3) (2009) 538–558.
[33] C. Roberts, Demonstratives as definite noun phrases, in: K. van Deemter, R. Kibble (Eds.), Information Sharing: Reference and Presupposition in Language Generation and Interpretation, CSLI Publications, Stanford, CA, 2002, pp. 89–106.
[34] N.J. Smith, R. Levy, The effect of word predictability on reading time is logarithmic, Cognition 128 (3) (2013) 302–319.
[35] T.P. Urbach, K.A. DeLong, M. Kutas, Quantifiers are incrementally interpreted in context: more than less, J. Mem. Lang. 83 (2015) 79–96.
[36] T.P. Urbach, M. Kutas, Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation, J. Mem. Lang. 63 (2) (2010) 158–179.
[37] C. Van Petten, Examining the N400 semantic context effect item-by-item: relationship to corpus-based measures of word co-occurrence, Int. J. Psychophysiol. 94 (3) (2014) 407–419.
[38] C. Van Petten, B.J. Luka, Prediction during language comprehension: benefits, costs, and ERP components, Int. J. Psychophysiol. 83 (2) (2012) 176–190.
[39] C., Dudschig, I.G., Mackenzie, J., Strozyk, B., Kaup, H., Leuthold. The sounds of sentences: Differentiating the influence of physical sound, sound imagery, and linguistically implied sounds on physical sound processing. Cognitive, Affective, and Behavioral Neuroscience, (in press).
[40] H. Nolan, R. Whelan, R.B. Reilly, FASTER: fully automated statistical thresholding for EEG artifact rejection, J. Neurosci. Methods 192 (1) (2010) 152–162.
[41] S. Prasada, A. Salajegheh, A. Bowles, D. Poeppel, Characterising kinds and instances of kinds: ERP reflections, Lang. Cognitive Proc. 23 (2) (2008) 226–240.