Brain and Cognition 103 (2016) 38–49

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Is there a difference between stripy journeys and stripy ladybirds? The N400 response to semantic and world-knowledge violations during sentence processing

Carolin Dudschig *, Claudia Maienborn, Barbara Kaup

University of Tübingen, Germany

ARTICLE INFO

Article history:
Received 16 June 2015
Revised 8 December 2015
Accepted 4 January 2016

Keywords:
Language comprehension
Semantic violations
World-knowledge violations
N400
EEG/ERP

ABSTRACT

The distinction between linguistic and non-linguistic knowledge is particularly relevant because it is related to the principle of compositionality during sentence comprehension. Hagoort, Hald, Bastiaansen, and Petersson (2004) challenged the distinction between linguistic and non-linguistic knowledge. Here, we investigate how linguistic and non-linguistic violations are processed in a setting adapted from Hagoort et al., whilst in contrast to Hagoort, keeping the critical word identical. In line with the findings by Hagoort et al., our results showed largest N400 amplitudes for semantic violations ('Journeys are stripy'), followed by non-linguistic world-knowledge violations ('Ladybirds are stripy') and contingent sentences ('Trousers are stripy'), and finally by correct sentences ('Zebras are stripy'). Traditional fractional area and relative criterion measures of peak and onset latencies showed no effect of violation type. Interestingly, the semantic violation condition crossed a fixed criterion earlier than the word-knowledge violation condition. In conclusion, our data suggests that the question regarding the distinction between linguistic- and non-linguistic knowledge in terms of language integration remains open. Implications for future studies addressing the difference between linguistic and non-linguistic knowledge are discussed.

1. Introduction

Words are the building blocks of language. However, what makes human language so fascinating is the ability to express an unlimited number of thoughts and ideas. Language allows us to communicate about almost everything, for example, about things we have experienced in the past, things we would like to happen in the future, but also about things that will never happen or do not even exist in this world. In order to do so, we use various syntactic rules that combine the words into sentences. As these rules typically work on the sentence level, the sentence has often been regarded as the core unit of language production and comprehension. Semanticists widely agree that some version of Frege's (1892) principle of compositionality must hold (e.g. Pagin & Westerståhl, 2011), according to which the meaning of a complex expression is a function of the meanings of its parts and the ways they are syntactically combined. However, other sources of knowledge that do not purely derive from our linguistic knowledge about the meaning of words or syntactic structures are also accessed during language comprehension. One prominent source of this type of non-linguistic knowledge is our knowledge about the world (Bierwisch, 2007; Hobbs, 2011; Isberner & Richter, 2014; Jackendoff, 2002; Jackendoff, 2003; Lang & Maienborn, 2011; Marques, Canessa, & Cappa, 2009; Richter, Schroeder, & Wöhrmann, 2009). In the past decades a lively discussion has evolved regarding the question whether linguistic and non-linguistic sources of knowledge are accessed and integrated separately during language comprehension or whether they are processed in an all-in-one step.

Models that assume the principle of compositionality typically suggest that during meaning composition the semantics of an expression is derived purely on the basis of our linguistic knowledge (knowledge about word meaning, syntactic structure, etc.), and is thus separate from the sentence's truth-value with regard to our knowledge about the world (Hagoort & van Berkum, 2007). For example, a sentence such as *Ladybirds are stripy* allows building a coherent semantic interpretation and is unproblematic when only taking into account our linguistic knowledge about the words in the sentence and its syntactic structure. However, if we additionally take into account our knowledge about the world, then we must conclude that the sentence is not true, as ladybirds

* 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.bandc.2016.01.001
0278-2626/

in our world typically are dotted and not stripy. In contrast, for a sentence such as *Journeys are stripy*, a violation is already indicated when only taking into account our linguistic knowledge. Specifically, the adjectival predicate 'stripy' demands an argument of type 'physical object', and thus cannot be combined with an event-denoting expression (i.e. *journey*) in order to derive a coherent semantic interpretation. That is, violation of a predicates's selectional restrictions leads to semantic anomaly, which is to be distinguished from falsity (world knowledge violation); see, e.g. Asher (2011).

One straightforward possibility to transfer the Fregean principle of compositionality into a processing account is thus to assume a two-step model of language comprehension, as outlined by Hagoort and van Berkum (2007): In a first step, the meaning of a sentence is constructed locally, by taking into account the meaning of the individual words and their syntactic combination. Other types of knowledge are integrated in a second step of processing. These other types of global pragmatic knowledge include the comprehender's general world-knowledge, knowledge resulting from the prior discourse and knowledge about the context of the utterance such as knowledge about the speaker's identity and gestures.

However, the empirical basis for such a model of sentence comprehension is rather controversial. Lattner and Friederici (2003) showed in an electrophysiological (EEG) study that non-linguistic information – specifically the information about a speaker's voice (male vs. female) and about gender stereotypes – was integrated rather late during sentence comprehension and occurred after the initial semantic analysis of a sentence had been completed. This finding seems to fit well with a two-step model of language comprehension. In contrast, recent studies question the assumptions regarding a separated integration of linguistic and non-linguistic knowledge during language understanding and instead suggest a one-step model of comprehension (for a review see: Hagoort & van Berkum, 2007). Indeed, manifold studies succeeded in showing that knowledge from various information sources are taken into account as early as sentence-based linguistic information during sentence comprehension. A key study by Hagoort et al. (2004) investigated whether the brain reflects differences between processing world-knowledge and semantic violations during sentence understanding. In order to differentiate between a one-step and a two-step model of sentence comprehension, EEG recordings were the central measurements in this study. EEG recordings typically allows one to measure real-time brain activity, within a resolution of milliseconds, and consequently are often used to investigate the precise time-course of cognitive processes (e.g., Dudschig & Jentzsch, 2008). In Hagoort et al.'s study, participants read sentences whose first conjunct expressed either a true statement such as *The Dutch trains are yellow (and very crowded)*, a world-knowledge violation such as *The Dutch trains are white (and very crowded)* or a semantic violation such as *The Dutch trains are sour (and very crowded)*. Thus, in order to create linguistic violations, the authors use the selection restriction criterion – whereby they violate the selection criterion that 'a predicate requires an argument whose semantic features match that of its predicate' (*p*. 439) – (see also Warren & McConnell, 2007; for alternatives, see, Pylkkänen, Oliveri, & Smart, 2009). For each sentence the N400 was measured relative to critical word onset (i.e. *yellow*, *white* and *sour*). The results showed that there was no difference in the N400 onset or peak latency between the semantic and the world-knowledge violations. Also the amplitude differences were rather small – however significant – when comparing the semantic and the world-knowledge violation conditions. The authors concluded that world-knowledge violations and semantic violations are integrated in parallel during language comprehension, in contrast to what is proposed by two-step-models of language comprehension. This study is often referenced as a milestone within psycholinguistic research by showing that empirically no difference between integrating linguistic and non-linguistic violations can be shown. In the following years, this work has been very influential and built the basis for many psychological theories of language understanding suggesting that all types of knowledge come down to one source (e.g., Filik, 2008; Matsuki et al., 2011; McRae & Matsuki, 2009; Nieuwland & Kuperberg, 2008). Despite challenges from linguistic research suggesting that there are serious grounds for the distinction between linguistic and non-linguistic knowledge, the N400 results from Hagoort and colleagues have been widely interpreted as indicating that there is no evidence from empirical investigations that suggest a processing or integration difference.

Given the importance of the results by Hagoort et al. (2004) for psycholinguistic research, we resume the question whether there are any indications that comprehenders integrate semantic knowledge and world knowledge in separate time steps during sentence comprehension. Specifically, we investigated the brain's responses to different violation types within sentences. As in the study by Hagoort and colleagues we used the N400-complex (Kutas & Hillyard, 1980) as our main measurement and implemented a conceptual replication of the original study with minor – but potentially important - changes in the experimental setup (for discussion on the value of such replication studies see: Schmidt, 2009). Surprisingly, it is still widely debated what mechanisms underlie the N400-complex (for a review see Lau, Phillips, & Poeppel, 2008). An *integration view* proposes that the N400 reflects semantic integration processes, whereby the N400 amplitude is determined by the ease with which the critical word can be integrated in the specific context (e.g., Brown & Hagoort, 1993). Alternatively, there is the *lexical view* of the N400 suggesting that the N400 reflects non-combinatorial processes and is determined only by the ease of access to long-term memory representations that are associated with one particular word in a specific context (for a review see Kutas & Federmeier, 2000). Independent of the exact mechanisms underlying the N400 effect, comprehensive work has investigated the specific parameters influencing the N400 complex. For example, the N400 is highly sensitive to word frequency (e.g., Kutas & Hillyard, 1984) and semantic distance (i.e. word co-occurrence; Van Petten, 2014). In the current study, we therefore used materials that allowed controlling for these issues. In line with Hagoort et al.'s study, we used correct sentences (*Zebras are stripy*), semantically violated sentences (*Journeys are stripy*), and sentences violated according to general world knowledge (*Ladybirds are stripy*). Importantly, in all our conditions, the critical words were kept identical. This allowed us to rule out word-based effects on the N400 complex and thus decreases the general N400 variance between the conditions, resulting in increased power to find between-condition differences (see Barber, Doñamayor, Kutas, & Münte, 2010; Davidson, Hanulikova, & Indefrey, 2012; Filik, Leuthold, Moxey, & Sanford, 2011; Hinojosa, Moreno, Casado, Muñoz, & Pozo, 2005). We also used an increased number of correct filler trials (50%) in order to avoid a strong bias of the participants in expecting a violated sentence and developing unnatural reading and comprehension strategies (see Kretzschmar, Bornkessel-Schlesewsky, & Schlesewsky, 2009). Additionally, we constructed contingent sentences – such as *Trousers are stripy* – containing a statement that is true for some members of the mentioned category but false for others, in order to avoid an instant and automated truth-value judgment triggered merely by the choice of the sentences used here. If there is a difference between the integration of linguistic and non-linguistic knowledge as predicted by two-step models of language comprehension, we expect differences in the N400 amplitudes and/or latencies between the semantic-violation and the world-knowledge violation condition. In contrast, if linguistic and non-linguistic sources of knowledge are integrated simultaneously during comprehension, no such differences should be observed.

As mentioned above, the central conclusions drawn from Hagoort et al.'s (2004) paper are based on the timing of the N400 complex. In the Hagoort study the onset latencies were calculated by testing when the violation conditions started to differ from each other. The difference waveform between the semantic and the world-knowledge violation condition was calculated and it was tested at what point in time this difference waveform significantly started to deviate from zero. The present study aimed at investigating the difference between semantic and world-knowledge violations with three main differences to the original study. First, the critical word was kept identical, in order to reduce word-based variance that might hide subtle differences between violation conditions. Second, additional correct and contingent sentences were added to the material in order to avoid biased reading strategies. Finally, in addition to traditional methods, recently established methods for detecting significant clusters in the EEG signal and dealing with the multiple comparisons problem in biological data (Maris & Oostenveld, 2007) were used in order to draw conclusions regarding the integration of world- versus semantic knowledge during sentence comprehension

2. Method

2.1. Participants

Thirty right-handed participants were included in the reported analysis (M_{age} = 22.70, SD_{age} = 2.72, range: 19–27, 8 male). All participants were native German speakers with normal or corrected to normal vision. Overall 35 participants were initially tested with five participants being removed due to excessive EEG artifacts resulting in a limited number of trials remaining for data processing (<60% of the trials accepted after preprocessing the data).

2.2. Material

Overall 360 German sentences were used in this experiment. 160 critical sentences and 180 filler sentences were constructed. Additionally 20 sentences were constructed for the practice trials. The critical sentences were divided into four subgroups. 40 sentences were correct sentences – true condition (e.g. '*Zebras are stripy*'). 40 sentences were constructed that were true for some members of the mentioned category – contingent condition (e.g. '*Trousers are stripy*'). 40 sentences were semantically violated – semantic condition (e.g. '*Journeys are stripy*'). Finally, another set of 40 sentences contained a world-knowledge violation – world-knowledge condition (e.g. '*Ladybirds are stripy*'). All sentences consisted of a noun, the copula *sein* ('to be') and an adjective (see Appendix, Table 1 for example sentences). The filler sentences were constructed similarly to the critical sentences and contained only true statements (e.g., '*Stones are hard*') in order to minimize the violation/true ratio in the experiment. The words were displayed in black on a light gray background. The nouns used in the experimental sentences were controlled for frequency with the Leipziger Wortschatzportal, $F(3,156) = 1.74$, $p = .16$ ($M_{Correct}$ = 1895, $M_{Contingent}$ = 5364, $M_{Semantic}$ = 3392, $M_{WorldKowledge}$ = 2787). The critical words were the adjectives; these were identical across all four conditions and each participant saw each adjective once in each condition. As a measure of word association between the noun and the adjective of each sentence, a latent semantic analysis (LSA; Landauer, Foltz, & Laham, 1998) was conducted using the LSAfun R-package (Günther, Dudschig, & Kaup, 2014) and previously established semantic spaces based on lemmatized newspaper and blog corpora (Günther, Dudschig, & Kaup, 2015). There was no main effect of LSA similarity between the four conditions, $F(3,156) = 1.44$, $p = .23$ ($M_{Correct}$ = 0.08, $M_{Contingent}$ = 0.06, $M_{Semantic}$ = 0.04, $M_{WorldKowledge}$ = 0.06). Given the rather small associations resulting from the LSA analyses an additional word association or predictability test was conducted. Specifically, as a type of cloze measurement, a sentence completion was conducted with twelve naïve German native speakers (M_{age} = 24.42, 10 female, all right-handed) that did not take part in the actual experiment. Participants were provided with the beginning of the sentence and had to provide five adjectives that came to their mind to end the sentence (Block & Baldwin, 2010; Bloom & Fischler, 1980; Taylor, 1953). If participants provided the correct adjective – according to our experimental sentences – this was counted as a hit, if they provided non-matching adjectives this was counted as a miss. Subsequently, the percentage of hits was calculated for each sentence. An ANOVA was conducted showing a clear effect of experimental condition in this cloze-measurement, $F(3,156)$ = 61.96, $p < .001$. Between conditions comparison t-tests showed that the correct sentences were more often completed with the according final word than contingent sentences, $t(78) = 5.60$, $p < .001$ ($M_{Correct}$ = 59.79%, $M_{Contingent}$ = 19.79%). The contingent condition also differed from the world-knowledge ($M_{WorldKnowledge}$ = 0.21%), $t(78) = 4.32$, $p < .001$, and the semantic violation condition ($M_{Semantic}$ = 0.00%), $t(78) = 4.38$, $p < .001$). Most importantly, the semantic and the world knowledge condition did not differ significantly, $t(78) = -1.00$, $p = .32$.

2.3. Procedure and design

The experiment took place in a sound-attenuated booth. The experimental procedure was controlled by E-Prime (Psychology Software Tools, Inc., Version E-Prime 2.0.10.92; Schneider, Eschman, & Zuccolotto, 2002). Each trial started with the presentation of a fixation-cross displayed in the center of the screen for 1500 ms. The words were presented successively in the center of the screen. Each word was displayed for 300 ms followed by a 300 ms blank screen. The last word was followed by a 1000 ms blank screen. In approximately 8% of the trials, comprehension questions were asked. Participants had to indicate with a button press (n or v key) whether a statement (e.g., '*Stones are solid material*') is correct or not according to the preceding sentence (e.g. '*Stones are hard*'). These trials were added to keep the participants' attention to the sentences. The experiment consisted of one practice block and four experimental blocks. Each experimental block contained 40 critical sentences, ten of each condition, and 45 filler sentences. Thereby it was ensured that critical words did not repeat within a block, in order to avoid direct word repetitions and that each condition was equally likely to be presented by the first, the second, the third or the fourth instance of the critical word. The trial order within each block was randomized for each participant. Block order was varied between participants (version 1: 1-2-3-4, version 2: 4-3-2-1, version 3: 3-1-4-2, version 4: 2-4-1-3).

3. Electrophysiological measures

3.1. EEG recording

EEG activity was recorded continuously from 70 electrodes using a BioSemi EEG system. The location of the electrodes are as follows: 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, P9, PO3, PO7, O1) and the homologue electrodes in the right hemisphere. Two additional electrodes (Common Mode Sense [CMS] active electrode and Drive Right Leg [DRL] passive electrode) were used as reference and ground electrodes, respectively (cf., www.biosemi/faq/cms&drl.htm). Vertical electroocular

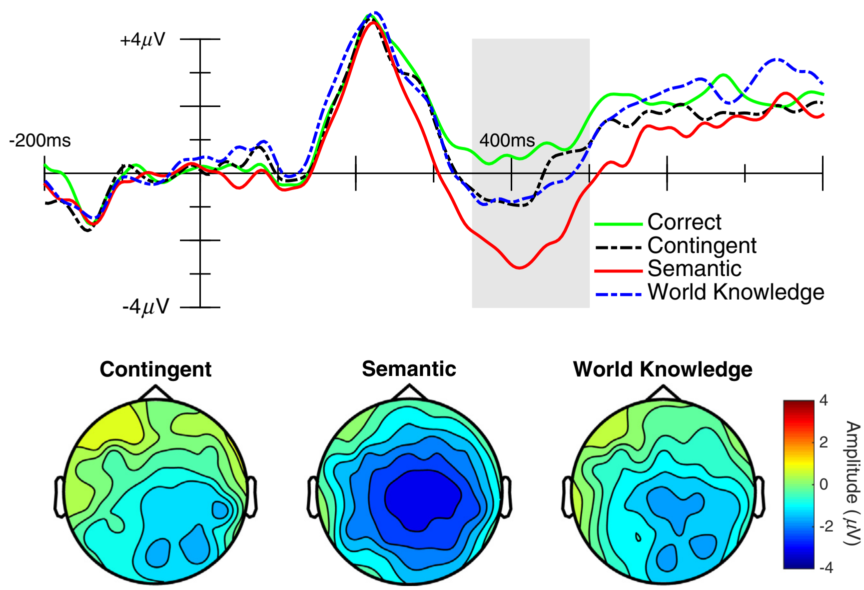

Fig. 1. Top panel: Event-Related-Potentials (ERPs) for the four conditions – correct, contingent, world-knowledge and semantic violation calculated as the mean from N400 electrodes as indicated by the topography (C1, Cz, C2, CP1, CPz, CP2, P1, Pz & P2). Bottom Panel: Topographic distributions of the N400 complex calculated as the difference between each violation condition and the correct condition in the 350–500 ms time interval. Early differences in the ERPs between the four conditions in the interval before the N400 measurement did not reach significance, there was also no significant difference between the four conditions in late positivities following the N400 interval up to 1600 ms.

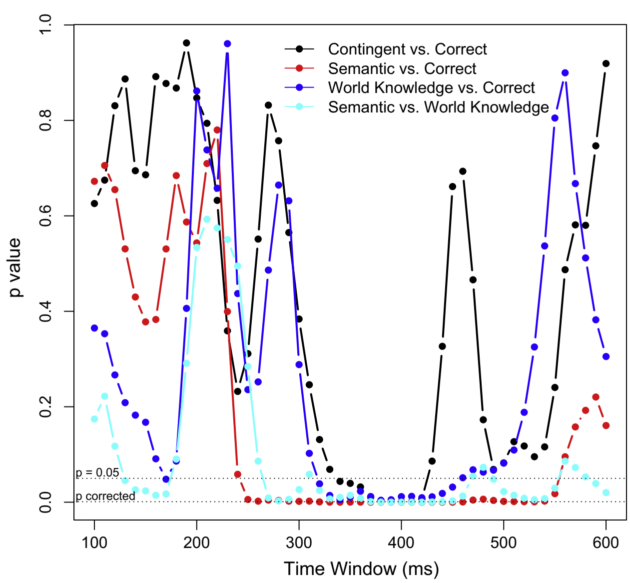

Fig. 2. The results of the running *t*-tests comparing each violation condition against the correct condition and additionally the comparison of the semantic condition against the world-knowledge condition are displayed. Both the non-corrected significance level ($p = .05$) and the Bonferroni corrected significance level are highlighted in the figure.

(vEOG) and horizontal EOG (hEOG) waveforms were calculated off-line as follows: vEOG(t) = Fp1(t) minus IO1(t) and hEOG(t) = F9(t) minus F10(t). The sampling rate for the EEG and electrooculogram (EOG) recordings was 256 Hz.

3.2. EEG analysis

All ERP analysis was performed using available MATLAB toolboxes (EEGLAB: Delorme & Makeig, 2004; FieldTrip: Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom MATLAB scripts. The analysis epoch started 1000 ms prior to the onset of the critical word and lasted 3 s. Off-line, all EEG channels were recalculated to an average mastoid reference and high-pass filtered (Butterworth IIR: 0.1 Hz, 36 dB/oct). Next, using a procedure similar to that described by Nolan, Whelan, and Reilly (2010), (ocular) artifacts were removed and EEG data were corrected. This procedure included a number of successive steps as described next. A predefined *z*-score threshold of ±3 was used to identify outliers relating to channels, epochs, independent components, and single-channels in single-epochs. In the first step, epochs containing extreme values in single electrodes (e.g., amplifier blockings, values larger ±1000 μV in any electrode) were removed, as were trials containing values exceeding ±75 μV in multiple electrodes that were not related to protypical eye movement artifacts. 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" EEG data set was subjected to a spatial independent components analysis (ICA) based on the infomax algorithm (Bell & Sejnowski, 1995; Jung et al., 2000). ICA components representing ocular activity (blinks and horizontal eye movements) were automatically identified using *z*-scored measures of the absolute correlation between the ICA component and the recorded hEOG and vEOG activity, respectively, and confirmed by visual inspection before being removed from the EEG data set. Fourthly, previously removed noisy channels were interpolated in the ICA-cleaned EEG data set using the average EEG activity of adjacent uncontaminated channels within a specified distance (4 cm, ~3–4 neighbors per electrode) in order to ensure a full electrode array for each participant. Approximately 1.3 electrodes were interpolated per participants across the whole data set. Following this procedure, there remained on average 147.73 trials per participant (out of 160; 92.33%). For artifact-free trials, the signal at each electrode site was averaged separately for each experimental condition, low-pass filtered (Butterworth IIR: 30 Hz, 36 dB/oct), and aligned to a 100-ms baseline prior to the onset of the critical word. Analysis time windows were based on visual inspection of the grand average ERP data. The mean ERP amplitudes were

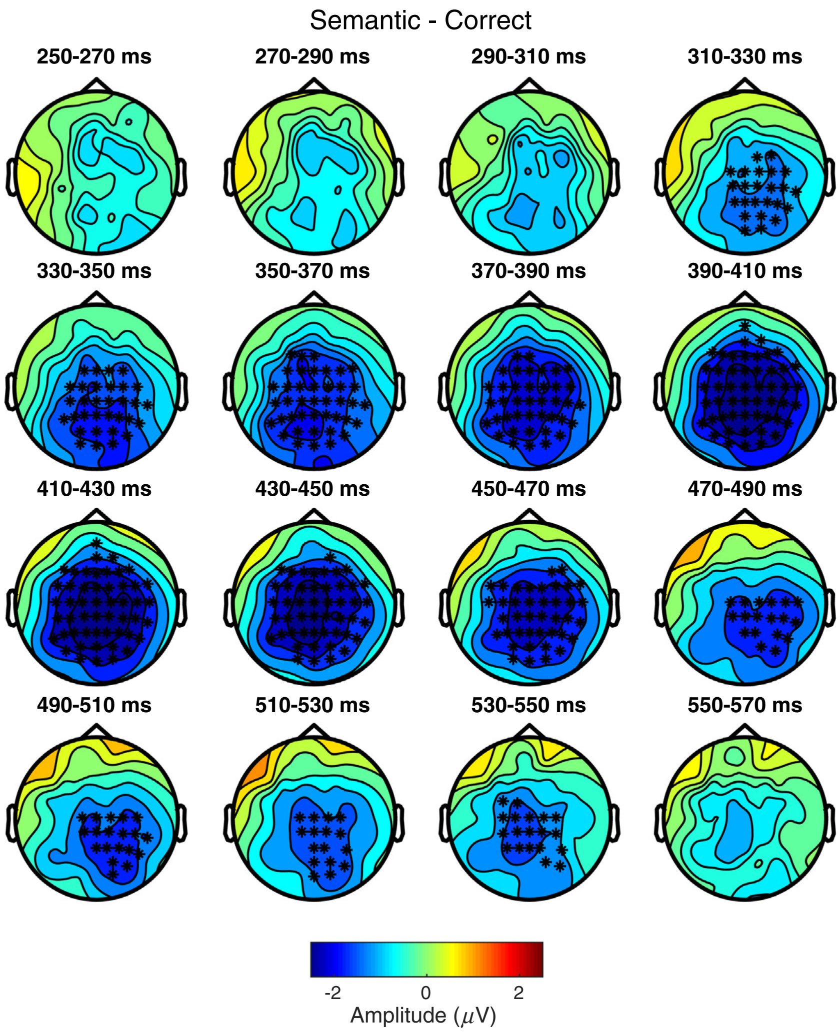

Fig. 3. Results of the permutations tests comparing the semantic condition to the correct condition. The black stars indicate the specific electrodes that showed significant differences between the semantic and the correct condition in the timing interval indicated above the single heads.

determined in the time window from 350 ms to 500 ms (N400 interval[1]) (Osterhout, Allen, & McLaughlin, 2002; Van Petten & Kutas, 1990). In order to further investigate the time course of significant differences between our correct, contingent, world-knowledge and semantic conditions, three approaches were adopted. First, in line with Hagoort and colleagues the time point at which the conditions started to deviate from each other and from the correct condition was analyzed. We therefore used newly developed cluster-based permutation tests across 64 scalp electrode sites within the time interval centered around 400 ms (200–600 ms post stimulus) with the constraint that at least two adjacent channels show a significant effect (see Maris & Oostenveld, 2007). These Montecarlo cluster-based permutation tests ($N = 1000$) are specifically designed to deal with multiple comparisons problems resulting from dense electrode layouts and multiple time-point testing. For determining significance in the cluster analysis we used two-sided t-tests with alpha set at 0.05 resulting in an alpha level of 0.025 at each tail. Second, we tested the N400 component peak latency using a 50% fractional area measure on the jackknifed waveforms as recommended by Kiesel, Miller, Jolicœur, and Brisson (2008) and also implemented a relative criterion method. Third, the time when a fixed criterion was crossed by each of the jackknifed waveforms from the violation conditions was compared. Where appropriate Greenhouse–Geisser corrected p-values for the repeated measures of variance are reported.

4. Results

All filler trials were excluded from data analysis. Several ANOVAs with the factor violation-type (correct, contingent, semantic vs. world-knowledge) were conducted in order to analyze the N400 mean amplitudes and latencies.

4.1. Amplitude

The ERP results are displayed in Fig. 1 (and Fig. 8, Appendix). The ERP analysis was conducted using the mean from central N400 electrodes (C1, Cz, C2, CP1, CPz, CP2, P1, Pz & P2). The ANOVA showed a main effect of violation-type, $F(3,87) = 16.27$, $p < .001$ ($\varepsilon = 0.90$). To further explore this main effect of violation type, single comparisons were conducted. There was a significant difference between the correct and the contingent sentences, $F(1,29) = 6.13$, $p < .05$, as well as between the correct and the world-knowledge violation sentences, $F(1,29) = 8.81$, $p < .01$ and the correct and the semantic violation sentences, $F(1,29) = 32.30$, $p < .001$. The contingent and world-knowledge violation sentences

[1] As some studies use the 300–500 ms post-stimulus time-interval for the N400, we also analyzed this time-interval, the results statistically fully mirrored the hereby reported results from the 350 ms to 500 ms time-interval. The same applied to standard Cz electrode analysis only.

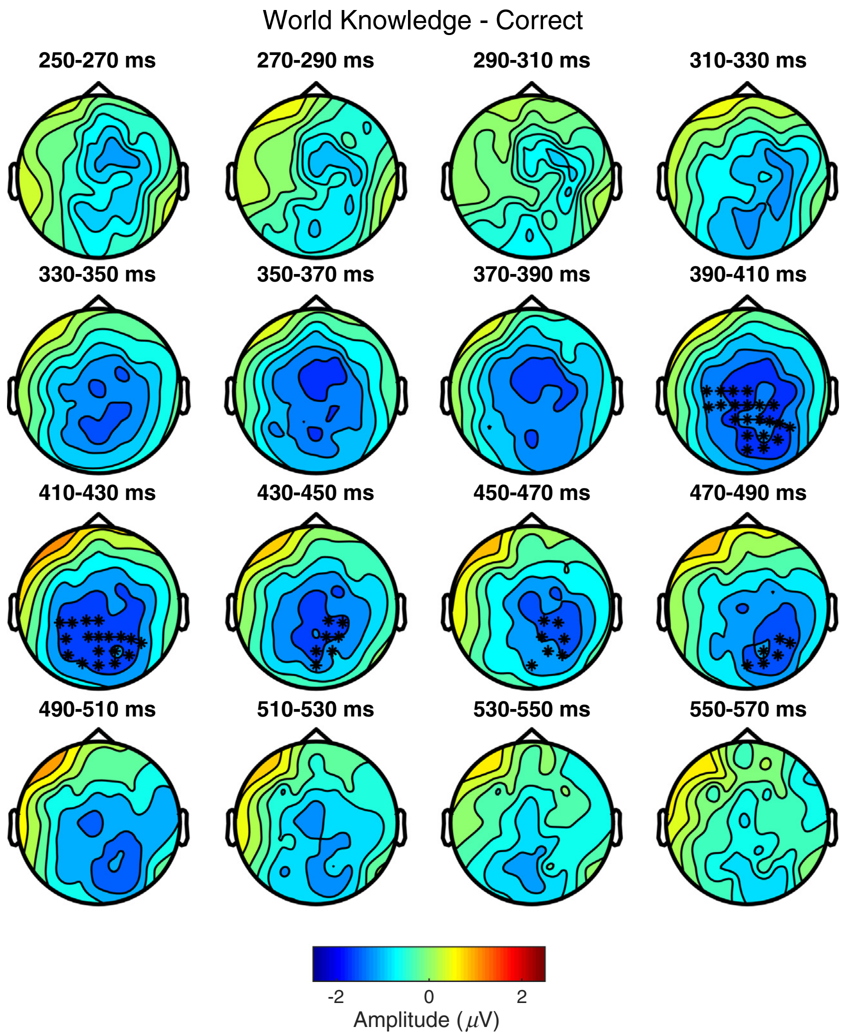

Fig. 4. Results of the permutations tests comparing the world-knowledge condition to the correct condition. The black stars indicate the specific electrodes that showed significant differences between the world-knowledge and the correct condition in the timing interval indicated above the single heads.

did not differ from each other, $F < 1$. Finally, both the contingent and the world-knowledge violation condition differed significantly from the semantic violation condition, $F(1,29) = 24.77$, $p < .001$ and $F(1,29) = 14.59$, $p < .001$, respectively. These findings regarding an amplitude difference between the semantic and the world-knowledge violation condition were in line with the original study by Hagoort et al. (2004).

4.2. Running t-tests

In a first step we implemented the identical analysis as used in the original paper (see supplemental material in Hagoort et al., 2004) to investigate the time course of the N400 effect. Running *t*-tests were conducted using 25 ms bins moving through the data by shifting the bin in 10 ms steps starting at 100–600 ms after critical word onset. The results are displayed in Fig. 2. Most importantly, the semantic and the world-knowledge condition start to differentiate at 270 ms if looking in the N400 interval. After conservative Bonferroni-correction this difference was consistently significant from 380 ms onwards.

4.3. Permutation tests

We performed analyses using recently developed permutation tests (Maris & Oostenveld, 2007) in order to investigate the course of the violation effects. Specifically, we focused on the differences between the violation conditions and the correct condition. Therefore, we compared the correct sentences with the contingent, the world-knowledge and the semantic violation condition. The cluster based permutation test revealed a significant central-parietal (i.e. standard N400 topography) difference between the correct and the semantic condition between 310 and 550 ms relative to stimulus onset (see Fig. 3). In contrast, the difference between the correct and the world-knowledge condition emerged at 390 ms (see Fig. 4), and the difference between the contingent condition and the correct condition around 350 ms (see Fig. 5). Additionally, when comparing the semantic and the world-knowledge violation (see Fig. 6) it can be seen that condition differences started to reach significance around the 390 ms time point. The results with the significant clusters highlighted are also displayed in the according difference waveforms in Fig. 7. In direct comparison to Hagoort et al. (2004) these analyses show a difference between the semantic and the world-knowledge condition emerging around 100 ms earlier than in the original study.

4.4. Peak and onset latencies

We also analyzed the N400 peak and onset latencies in the time window between 300 ms and 500 ms on the electrodes previously selected for the N400 amplitude analysis. First, the peak latencies

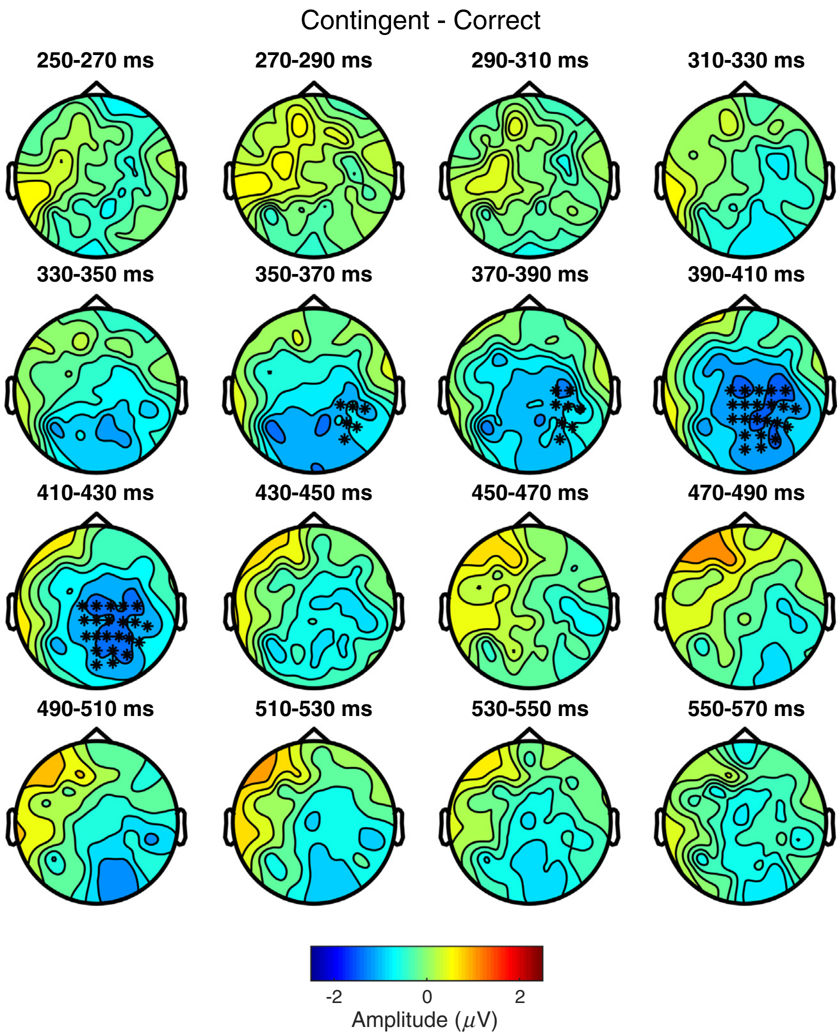

Fig. 5. Results of the permutations tests comparing the contingent condition to the correct condition. The black stars indicate the specific electrodes that showed significant differences between the contingent and the correct condition in the timing interval indicated above the single heads.

were determined by the fractional mean as suggested by Kiesel et al. (2008) on the jackknifed data with a 50% fractional area criterion in the 300–500 ms interval – that was revealed as the interval of interest by the permutation tests performed (see above). The boundary for the area analysis was set to 1.5 μV. p-Values were corrected according to Ulrich and Miller (2001) with $F_c = F/(n-1)^2$. The analysis showed no effect of peak latency ($F < 1$, $M_{Correct} = 394$ ms, $M_{Contingent} = 395$ ms, $M_{Semantic} = 407$ ms, $M_{WorldKnowledge} = 402$ ms). Second, as analysis of onset differences the relative criterion measure was calculated analogously to the fractional area calculations, determining the time in point where 50% of the peak amplitude was exceeded in each condition. Again, there was no effect in timing ($F < 1$; $M_{Correct} = 330$ ms, $M_{Contingent} = 325$ ms, $M_{Semantic} = 322$ ms, $M_{WorldKnowledge} = 321$ ms). These types of onset analyses were not conducted in the original study by Hagoort et al. (2004). However, the results are fully in line with the interpretation of the original study. Finally, a new way of N400 onset analysis was conducted. Here we determined the time point when each of the four conditions reached a fixed 2 μV criterion starting from the peak preceding the N400. This type of analysis was suggested (personal communication Jeff Miller) given the large differences in the N400 amplitude between the four conditions, a situation that is typically not ideal for determining component onset by the relative criterion method (see also Kiesel et al., 2008). Additionally, the clear alignment of the four conditions on the peak before the N400 encourages such an peak to fixed criterion analysis. After jackknife-corrections, the results showed a trend for the main effect condition, $F(3,87) = 2.55$, $p = .06$. Follow-up comparisons showed that the contingent condition did not differ from the correct condition ($F < 1$), the world-knowledge condition did not differ from the correct condition ($F < 1$), and the semantic condition did differ from the correct condition, $F(1,29) = 4.69$, $p < .05$. Most interestingly, the semantic condition also differed significantly from the world-knowledge condition with regard to the time point when the 2 μV criterion was reached, $F(1,29) = 4.22$, $p < .05$. Additionally, the semantic condition also differed from the contingent condition, $F(1,29) = 4.88$, $p < .05$. Taken together the different conducted analyses of N400 onset or peak latencies report mixed results: On the one side, the traditional methods (i.e. fractional area and relative criterion) taking the peak as the reference point, analyzing backwards show no difference regarding the N400 onset or peak latency. On the other side, a new exploratory method shows a difference between the time point when the semantic and the world-knowledge violation reach a fixed criterion starting from the peak before the N400. Probably, these results have to be carefully interpreted. However, they take into account the current situation with the rather large differences in amplitude. Moreover, in contrast to the permutation and the

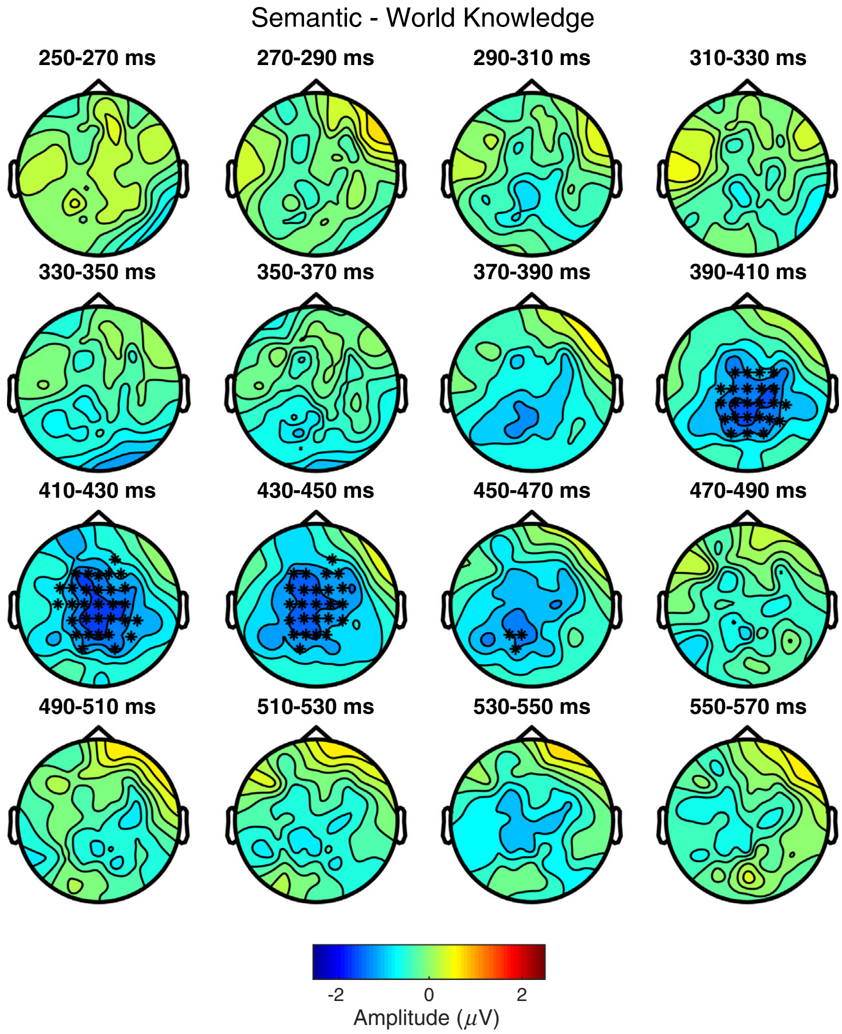

Fig. 6. Results of the permutations tests comparing the semantic condition to the world-knowledge condition. The black stars indicate the specific electrodes that showed significant differences between the semantic and the world-knowledge condition in the timing interval indicated above the single heads.

running *t*-tests they also focus on the analysis of a time point, rather than testing for amplitude differences.

5. Discussion

At present we are far from fully understanding what processes are involved during sentence comprehension. One important debate refers to the question of whether all of our knowledge is integrated in a one-step procedure during sentence comprehension, or whether linguistic knowledge – that is knowledge about word meanings and syntactic structures has priority during meaning comprehension, followed by other sources of knowledge, for example, our world-knowledge, knowledge about the speaker's background and other non-linguistic types of knowledge (e.g., speakers voice or accent). Hagoort et al. (2004) investigated whether semantic and world-knowledge violations result in different N400 onset latencies during sentence comprehension. They reported no timing difference regarding the integration of linguistic and non-linguistic knowledge and suggested that these results point towards a one-step model of language comprehension. To date this study is often cited as key evidence suggesting that linguistic and non-linguistic sources of knowledge are integrated simultaneously during language processing. Given the importance of these findings for all models of language comprehension we took this question up again and implemented an experimental setup and analysis methods that allow to shed new light on this discussion.

Our study was designed with direct reference to the original study by Hagoort et al. (2004). Thus, our study can be regarded as a conceptual replication of this very influential study with a few changes in the experimental setup and the analysis methods applied. There were two main differences to the original study. First, in contrast to Hagoort et al. (2004), we used an experimental setup where the critical word was kept identical between conditions. This allowed us to measure the N400 on the same word in each violation condition, which has the advantage of reducing word-based effects on the N400 complex. Second, we used recently developed nonparametric permutation-based analysis methods that at the same time are sensitive to picking up differences by taking into account biophysical constraints in the testing procedure but also are able to deal with the multiple comparison problem (see Maris & Oostenveld, 2007). When analyzing standard measures for component onset – that is the fractional area under the N400 curve and the relative-criterion peak latency measure – no differences in onset latencies were observed. This finding is in line with the conclusions drawn in the Hagoort et al. study and can be interpreted as pointing towards a one-step model of language comprehension. Interestingly, in the original study a different measure for onset of the semantic and the world-knowledge violation

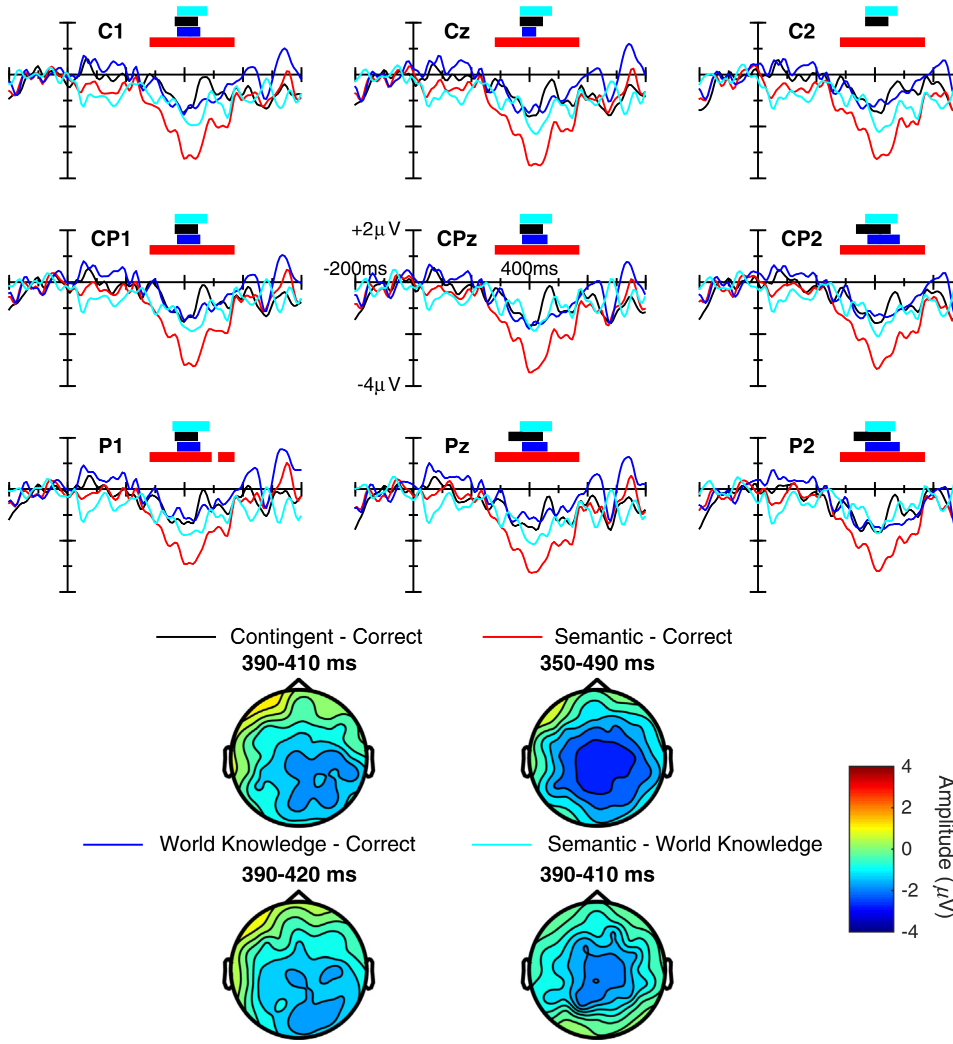

Fig. 7. Difference waveforms for each violation condition ('violation condition' – 'correct condition') and for the direct comparison of the semantic and the world-knowledge condition for the subset of central electrodes selected for the N400 analysis. The bars represent the time intervals of significant difference between each violation condition and the correct condition and the topo plots the according scalp distribution for the significant clusters.

was calculated. Specifically, Hagoort and colleagues compared the violations conditions by moving a 25 ms bin through the dataset by means of 10 ms shift and comparing the amplitude means of the conditions. Here, no difference between the violation conditions was observed until 480 ms after critical word onset, this result was interpreted as a lack of difference in component onset. We implemented the identical analysis and found that the difference between the semantic and world-knowledge violations occurred around 270 ms after word onset (380 ms after conservative Bonferroni correction). These results regarding a slightly earlier differentiation between the critical conditions – compared to the original study – were confirmed when using cluster-based permutation tests. Taken together both of these analyses indicated that there is a clear amplitude difference between the semantic and the world-knowledge violation condition. Critically, standard analyses of onset or peak timing (as reported above) are typically designed for situations where there is no large difference in amplitude between the conditions of interest (Kiesel et al., 2008). Thus, we performed an additional timing analysis – starting at the peak preceding the N400 – that should be less sensitive to amplitude differences. This analysis clearly indicated that the semantic and the world-knowledge condition differed with regard to the time point when they crossed a fixed criterion. One might think that these results suggest that there are specific timing differences between the integration of linguistic and non-linguistic sources of knowledge during sentence comprehension. However, there are several issues that need to be discussed before any far-reaching conclusions can be drawn. When taking all types of timing analysis together, we are now left with a range of partly opposing – and partly difficult to interpret results. In the following sections we will discuss these results and their contributions with regard to the main question of a one-step versus a two-step model of sentence comprehension in detail.

As mentioned in the introduction, the debate regarding the exact mechanisms reflected in the N400 complex is ongoing (for a review see Lau et al., 2008). The *lexical view* suggests that the N400 reflects non-combinatorial processes, whilst the *integration view* suggests that sentence integration processes are a major contribution to the N400. Strictly speaking, before interpreting any type of N400 result as evidence for or against the assumption that linguistic and non-linguistic knowledge sources are taken into account successively during sentence comprehension, there should be some consensus on whether the N400 really does reflect sentence-based integration processes in such an experimental setup. Critically, it is not fully understood what the N400 complex reflects in general, and more specifically, it is also an open issue which subpart of the complex reflects which process. What is the critical time point where a potential integration process takes place? Is it the beginning when the N400 starts to deviate from the correct condition or is it at the time-point of the N400 peak? The results of the current study therefore also contribute to the ongoing discussion about the processes underlying the N400 complex, and its sensitivity – with regard to amplitude, peak latency and onset latency - to specific types of violation.

Another – but related – issue that needs to be resolved is the interpretation regarding specific markers for the N400 analysis. As described above all our conditions show similar peak latencies. Indeed, it just became of interest within N400 research that the N400-complex is notably stable regarding its latency. The N400 only responds with amplitude changes rather than with changes in peak latency to different types of violations or the violation's strength (Federmeier & Laszlo, 2009; Kutas & Federmeier, 2011). However, this observation seems only recently to become a matter of discussion and a "fact whose theoretical significance we are just beginning to appreciate" (p. 625, Kutas & Federmeier, 2011). Thus, the N400 peak latency does not seem to be affected by experimental manipulations other than between subject variables such as age (Federmeier & Laszlo, 2009; Kutas & Federmeier, 2011). In line with this assumption, our results indicate that the N400 peak latency – determined by the fractional area or the relative-criterion technique – is extremely stable across violation conditions. In addition to these traditional timing analyses we performed a new way of analyzing the N400 time course. This peak to criterion analysis might be particularly interesting in situations as described in the current study – where the focus is on between-condition comparisons when there is a rather large amplitude difference. Specifically, we analyzed the ERP-waveforms starting at the peak preceding the N400 and tested when the conditions crossed a fixed criterion. Interestingly, in this analysis we do find some evidence suggesting that the semantic violation condition crossed this fixed criterion earlier than the world-knowledge violation condition, and the correct or contingent condition, whereas the

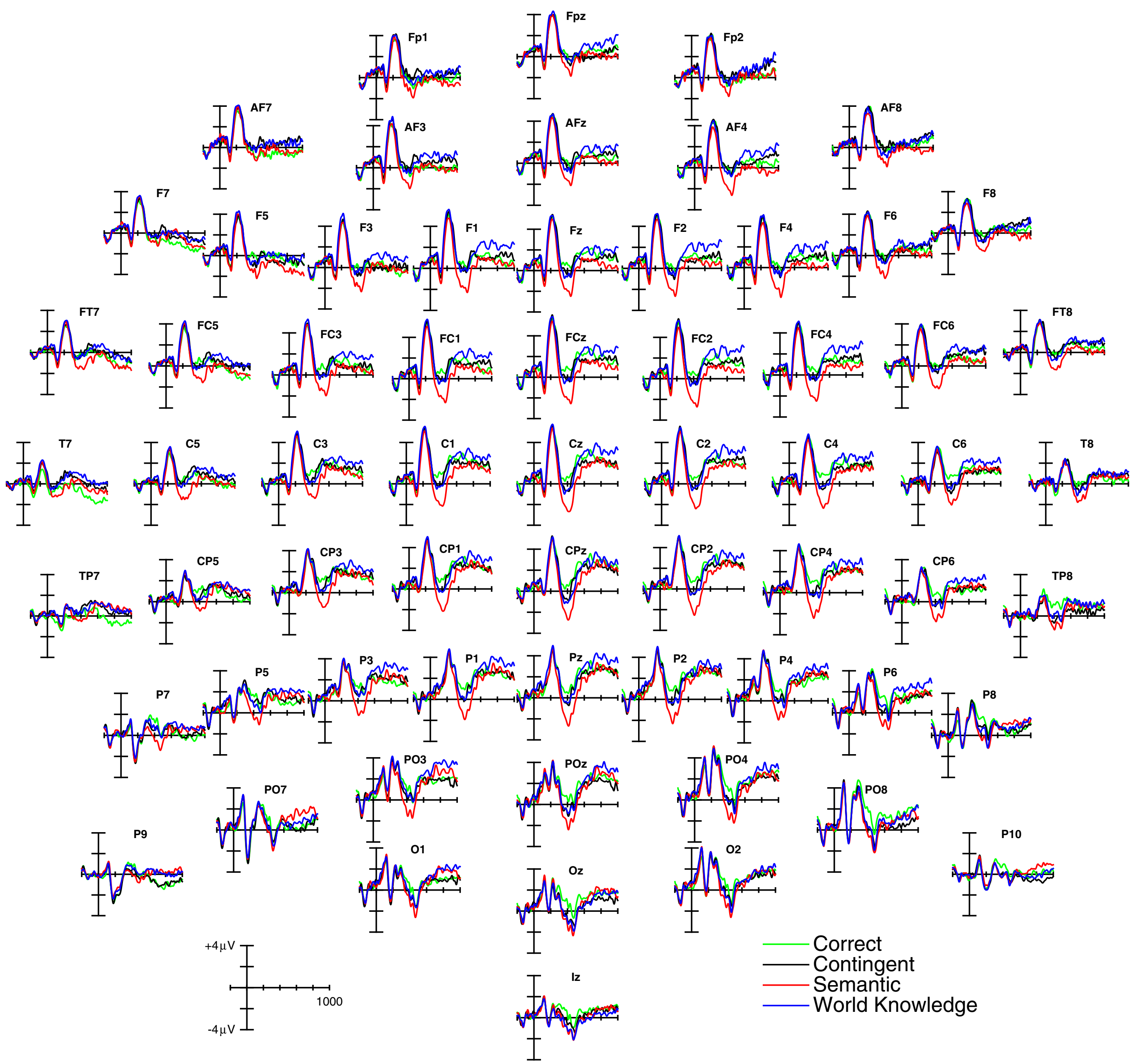

Fig. 8. Event-Related-Potentials (ERPs) for the four conditions – correct, contingent, world-knowledge and semantic violation for all electrodes separately.

other conditions did not differ significantly from each other. Thus, with regard to the starting point of the N400 complex, namely when it starts to deviate from the correct condition, one might argue that we did see a difference between the critical conditions in this type of analysis. One might take this as evidence for two-step models of language comprehension. However, in our view one has to be very careful with this interpretation. One reason making interpretation difficult is the finding that the traditional analysis of onset differences (fractional area and relative-criterion) clearly did not show any timing differences. Even if this new method might be the appropriate way of analysis for such situations where there also exists an amplitude difference, it has to be treated with care. As long as one does not fully understand which process underlies the N400 peak time and which process are involved when the N400 starts to develop, one cannot draw final conclusions regarding the cause of the observed differences. Still, in our study we used an experimental setup directly adapted from Hagoort et al. (2004), we had the same rationale for our analysis – that was the aim to investigate the time-point when the violation conditions started to differ from the correct condition – and in contrast to Hagoort and colleagues we performed more direct tests of timing. In these tests we did find a difference between the time when the semantic violation condition and the world-knowledge violation condition crossed a fixed criterion. We also found an amplitude difference between the semantic and the world-knowledge violation condition starting in a time interval 100–200 ms earlier than in the original study by Hagoort and colleagues. Thus, at the very least our results suggest that the question whether linguistic and non-linguistic knowledge are taken into account successively or simultaneously during sentence comprehension is far from settled and should be revisited in future studies (see also: Milburn, Warren & Dickey, in press; Paczynski & Kuperberg, 2012).

In order to move on towards answering these questions there are several routes that can be taken. First, instead of de-contextualized written sentences more real-life language processing should be investigated where participants have a context about what is communicated and thus might be better or worse in dealing with specific types of violations (see also: van Berkum, Hagoort, & Brown, 1999). Additionally, in order to fully understand whether the N400-effects observed in such studies as in the present study represent semantic versus world-knowledge violations other linguistic manipulations should be introduced. For example, one could use the negation operator in order to further differentiate the influence of word-based associations on the N400 measure in contrast to real truth-value violations. For example, the sentence *Ladybirds are not stripy* should result in smaller N400 effects than the sentence *Ladybirds are very stripy*, if indeed the N400 measure in such de-contextualized situations reflects sentential integration processes.

6. Conclusions

What makes language so fascinating is that it allows the combining of words by means of combinatorial rules into an unlimited number of sentences. A full understanding of how sentence meaning is constructed during comprehension would be a big step forward in understanding how our language system works. In the past it has been suggested that sentence comprehension takes place in a one-step fashion, whereby all information – linguistic and non-linguistic in nature – is integrated simultaneously during comprehension. Our study, that was specifically designed to answer such questions by looking at electrophysiological data and using newly developed analysis methods, – did show a difference between the processing of semantic and world-knowledge violations specifically with regard to amplitude. The analysis of latencies was more ambivalent. There was no difference in traditional analyses of onset latencies between the semantic and the world-knowledge violation. Only when considering the time point when a fixed criterion was crossed was there an indication that the semantic violation reached this criterion earlier than the world-knowledge violation, or the correct or contingent condition. These results indicate that the debate concerning one- versus two-step models of comprehension must be considered ongoing and should continue to be a central theme within psycholinguistic research.

Table 1
Example sentences used in this experiment in the correct, the contingent, the semantic and the world-knowledge violation conditions.

Example 1
Correct: Chilischoten sind scharf. (Chilies are hot)
Contingent: Suppen sind scharf. (Soups are hot)
World-Knowledge: Erdbeeren sind scharf. (Strawberries are hot)
Semantic: Wolken sind scharf. (Clouds are hot)
Example 2
Correct: Kreise sind rund. (Circles are round)
Contingent: Kuchen sind rund. (Cakes are round)
World-Knowledge: Würfel sind rund. (Dices are round)
Semantic: Geräusche sind rund. (Sounds are round)
Example 3
Correct: Säfte sind flüssig. (Juices are liquid)
Contingent: Medikamente sind flüssig. (Medicines are liquid)
World-Knowledge: Kekse sind flüssig. (Biscuits are liquid)
Semantic: Ängste sind flüssig. (Fears are liquid)
Example 4
Correct: Sprinter sind schnell. (Sprinters are fast)
Contingent: Cabrios sind schnell. (Convertibles are fast)
World-Knowledge: Schnecken sind schnell. (Schnecken sind schnell)
Semantic: Muster sind schnell. (Patterns are fast)

Acknowledgments

This project was supported by a Margarete-von-Wrangell Fellowship appointed to Carolin Dudschig (European Social Fund and the Ministry of Science, Research and the Arts Baden-Württemberg), by the Institutional Strategy of the University of Tübingen (German Research Foundation, ZUK 63), by the SFB833/B4 project of Barbara Kaup and Hartmut Leuthold, and by the SFB833/A1 project of Claudia Maienborn (German Research Foundation).

Appendix A

See Fig. 8 and Table 1.

References

Asher, N. (2011). *Lexical meaning in context. A web of words*. Cambridge: Cambridge University Press.
Barber, H. A., Doñamayor, N., Kutas, M., & Münte, T. (2010). Parafoveal N400 effect during sentence reading. *Neuroscience Letters, 479*, 152–156.
Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*(6), 1129–1159.
Bierwisch, M. (2007). Semantic form as interface. In A. Späth (Ed.), *Interfaces and interface conditions* (pp. 1–32). Berlin: de Gruyter.
Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods, 42*(3), 665–670.
Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition, 8*(6), 631–642.
Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5*(1), 34–44.
Davidson, D. J., Hanulikova, A., & Indefrey, P. (2012). Electrophysiological correlates of morphosyntactic integration in German phrasal context. *Language and Cognitive Processes, 27*, 288–311.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.

Dudschig, C., & Jentzsch, I. (2008). Locus of response slowing resulting from alternation-based processing interference. *Psychophysiology, 45*(5), 751–758.

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of Learning and Motivation, 51*, 1–44.

Filik, R. (2008). Contextual override of pragmatic anomalies: Evidence from eye movements. *Cognition, 106*(2), 1038–1046.

Filik, R., Leuthold, H., Moxey, L. M., & Sanford, A. J. (2011). Anaphoric reference to quantified antecedents: An event-related brain potential study. *Neuropsychologia, 49*, 3786–3794.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik, 100*, 25–50.

Günther, F., Dudschig, C., & Kaup, B. (2014). LSAfun – An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, 1–15.

Günther, F., Dudschig, C., & Kaup, B. (2015). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *The Quarterly Journal of Experimental Psychology.*, 1–28 (ahead-of-print).

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438–441.

Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 801–811.

Hinojosa, J. A., Moreno, E. M., Casado, P., Muñoz, F., & Pozo, M. A. (2005). Syntactic expectancy: An event-related potentials study. *Neuroscience Letters, 378*, 34–39.

Hobbs, J. (2011). Word meaning and world knowledge. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.). *Semantics: An international handbook of natural language meaning* (Vol. 1, pp. 740–761). Berlin: de Gruyter.

Isberner, M. B., & Richter, T. (2014). Does validation during language comprehension depend on an evaluative mindset? *Discourse Processes, 51*(1–2), 7–25.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Jackendoff, R. (2003). Précis of foundations of language: Brain, meaning, grammar, evolution. *Behavioral and Brain Sciences, 26*(06), 651–665.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(02), 163–178.

Kiesel, A., Miller, J., Jolicœur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology, 45*(2), 250–274.

Kretzschmar, F., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. *NeuroReport, 20*, 1613–1618.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470.

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.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2–3), 259–284.

Lang, E., & Maienborn, C. (2011). Two-level semantics: Semantic form and conceptual structure. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.). *Semantics: An international handbook of natural language meaning* (Vol. 1, pp. 709–740). Berlin: de Gruyter.

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing–evidence from event-related brain potentials. *Neuroscience Letters, 339*(3), 191–194.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920–933.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190.

Marques, J. F., Canessa, N., & Cappa, S. (2009). Neural differences in the processing of true and false sentences: Insights into the nature of 'truth' in language comprehension. *Cortex, 45*(6), 759–768.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 913–934.

McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass, 3*(6), 1417–1429.

Milburn, E. A., Warren, T., & Dickey M. W. (in press). World knowledge affects prediction as quickly as selectional restrictions do: Evidence from the visual world paradigm. *Language, Cognition and Neuroscience*.

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*(12), 1213–1218.

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods, 192*(1), 152–162.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 156869.

Osterhout, L., Allen, M., & McLaughlin, J. (2002). Words in the brain: Lexical determinants of word-induced brain activity. *Journal of Neurolinguistics, 15*, 171–187.

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language, 67*(4), 426–448.

Pagin, P., & Westerståhl, D. (2011). Compositionality. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.). *Semantics: An international handbook of natural language meaning* (Vol. 1, pp. 96–123). Berlin: de Gruyter.

Pylkkänen, L., Oliveri, B., & Smart, A. J. (2009). Semantics vs. world knowledge in prefrontal cortex. *Language and Cognitive Processes, 24*, 1313–1334.

Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology, 96*(3), 538.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*(2), 90–100.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime 1.0. Pittsburgh, PA: Psychological Software Tools.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly, 30*, 415–433.

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology, 38*(05), 816–827.

van Berkum, J. V., Hagoort, P., & Brown, C. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657–671.

Van Petten, C. (2014). Examining the N400 semantic context effect item-by-item: Relationship to corpus-based measures of word co-occurrence. *International Journal of Psychophysiology, 94*(3), 407–419.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequencyinevent-related brainpotentials. *Memory & Cognition, 18*(4), 380–393.

Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review, 14*, 770–775.