



Discourse-level semantic coherence influences beta oscillatory dynamics and the N400 during sentence comprehension

Ashley Glen Lewis, Jan-Mathijs Schoffelen, Christian Hoffmann, Marcel Bastiaansen & Herbert Schriefers

To cite this article: Ashley Glen Lewis, Jan-Mathijs Schoffelen, Christian Hoffmann, Marcel Bastiaansen & Herbert Schriefers (2016): Discourse-level semantic coherence influences beta oscillatory dynamics and the N400 during sentence comprehension, Language, Cognition and Neuroscience, DOI: [10.1080/23273798.2016.1211300](https://doi.org/10.1080/23273798.2016.1211300)

To link to this article: <http://dx.doi.org/10.1080/23273798.2016.1211300>



Published online: 20 Jul 2016.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Discourse-level semantic coherence influences beta oscillatory dynamics and the N400 during sentence comprehension

Ashley Glen Lewis^{a,b,c} , Jan-Mathijs Schoffelen^{b,c} , Christian Hoffmann^d, Marcel Bastiaansen^{b,e} and Herbert Schriefers^d

^aHaskins Laboratories, New Haven, CT, USA; ^bNeurobiology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; ^cCentre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; ^dCentre for Cognition, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; ^eAcademy for Leisure, NHTV University of Applied Sciences, Breda, The Netherlands

ABSTRACT

In this study, we used electroencephalography to investigate the influence of discourse-level semantic coherence on electrophysiological signatures of local sentence-level processing. Participants read groups of four sentences that could either form coherent stories or were semantically unrelated. For semantically coherent discourses compared to incoherent ones, the N400 was smaller at sentences 2–4, while the visual N1 was larger at the third and fourth sentences. Oscillatory activity in the beta frequency range (13–21 Hz) was higher for coherent discourses. We relate the N400 effect to a disruption of local sentence-level semantic processing when sentences are unrelated. Our beta findings can be tentatively related to disruption of local sentence-level syntactic processing, but it cannot be fully ruled out that they are instead (or also) related to disrupted local sentence-level semantic processing. We conclude that manipulating discourse-level semantic coherence does have an effect on oscillatory power related to local sentence-level processing.

ARTICLE HISTORY

Received 3 November 2015
Accepted 4 July 2016

KEYWORDS

Language comprehension;
discourse semantics; beta
oscillations; N400; EEG

Introduction

Language comprehension is an inherently dynamic process, with multiple sources of linguistic and non-linguistic information impinging upon the interpretation of just about any utterance or text. Exactly when during comprehension these different sources of information play a role is still an open question. Until recently the majority of research on language comprehension has been aimed at understanding the processing of individual words or single sentences in isolation. This has especially been the case for the investigation of oscillatory neural dynamics related to language processing. Yet everyday language use typically takes place within far richer contexts, and the information being conveyed goes beyond the meaning that can be decoded from single words or isolated sentences. Here, we take up the challenge, and investigate how discourse information affects oscillatory neural dynamics related to language comprehension.

Within the field of electroencephalography (EEG) research the analysis of event-related potentials/fields (ERPs/ERFs) has proven invaluable in exploring the timing of various types of linguistic processing (e.g. DeLong, Urbach, & Kutas, 2005; Friederici, 2002;

Hagoort & van Berkum, 2007). However, ERPs/ERFs provide only a glimpse into the rich spatio-spectro-temporal dynamics contained in the EEG/MEG (Magnetoencephalography) signal (Makeig, Debener, Onton, & Delorme, 2004). ERP/ERF analyses rely on averaging over trials and participants in order to improve the signal-to-noise ratio (SNR) to observe time- and phase-locked (to an event of interest) neural signatures in the EEG/MEG. However, not all neural activity related to an event is strongly phase-locked (evoked activity) to that event, and measuring non-phase-locked (induced), oscillatory activity (Tallon-Baudry & Bertrand, 1999) in the EEG/MEG can provide additional or complementary information about the underlying cognitive processing.

Neural synchronisation plays an important role in the coupling and uncoupling of functional brain networks (e.g. Pfurtscheller & Lopes da Silva, 1999; Singer, 1993; Varela, Lachaux, Rodriguez, & Martinerie, 2001). Functional networks are created by synchronous repetitive firing of populations of neurons, resulting in an increased probability that interacting neurons entrain one another in a rhythmic, frequency-specific manner (e.g. König & Schillen, 1991). This mechanism for the segregation of different types of information (represented in networks

firing synchronously at different frequencies) also supports the integration (or binding) of information distributed over distant neural populations (Gray, König, Engel, & Singer, 1989). Measuring frequency-specific oscillatory neural dynamics provides us with a window onto the dynamic coupling and uncoupling of such functional networks, and how this changes depending on the cognitive task.

A number of studies have shown a link between oscillatory activity in the beta frequency range (13–30 Hz) and manipulations of syntactic processing (see Lewis, Wang, & Bastiaansen, 2015 for review). Bastiaansen and Hagoort (2006) reported higher beta power for centre-embedded relative clauses compared to syntactically less complex right-branching relative clauses. Similarly, Weiss et al. (2005) report higher beta coherence between anterior and posterior electrodes for object-relative clauses compared to syntactically less complex subject-relative clauses. Meyer, Obleser, and Friederici (2013) found higher beta power for long- compared to short-distance subject–verb agreement dependencies at the point in the sentence where the dependency could be resolved. They argued that this was related to syntactic unification (Hagoort, 2005, 2013), since syntactic unification is likely more difficult in the case of long-distance dependencies. Finally, a number of studies have shown that beta power is higher for syntactically legal sentences compared to sentences containing a syntactic violation at the target word (Bastiaansen, Magyari, & Hagoort, 2010; Davidson & Indefrey, 2007; Kiehl, Meltzer, Moreno, Alain, & Bialystok, 2014; Kiehl, Panamsky, Links, & Meltzer, 2014). Together, these studies show that when syntactic unification becomes more difficult beta power increases, while disrupting syntactic unification leads to a relative decrease in beta power.

There is now also a large body of evidence linking oscillatory activity in the gamma frequency range (the findings are somewhat variable in terms of the exact frequency range, but all fall within the classical 30–100 Hz gamma band) to semantic processing (see Lewis et al., 2015 for review). Peña and Melloni (2012), for example, have shown that gamma power (55–75 Hz) increases while listening to sentences in one's own language, but not while listening to sentences in a language that one does not speak/understand (where semantic processing presumably does not take place). From a different perspective, van Berkum, Zwitserlood, Bastiaansen, Brown, and Hagoort (2004) reported higher gamma power for referentially correct words compared to words that had no referent or were referentially ambiguous (and hence the assignment of thematic roles was presumably disrupted). Finally, a number of studies have reported an

increase in gamma power at a target word for words that can be meaningfully integrated with a strongly constraining prior sentence context, but no gamma increase either when the target word results in a semantic violation, or when the sentence context is not strongly constraining (Hald, Bastiaansen, & Hagoort, 2006; Penolazzi, Angrilli, & Job, 2009; Rommers, Dijkstra, & Bastiaansen, 2013; Wang, Zhu, & Bastiaansen, 2012; Weiss & Mueller, 2003 for coherence instead of power). Wang, Zhu et al. (2012), for instance, compared sentences containing a high cloze probability (a measure of how well a particular target word completes a prior sentence context according to participants' offline judgements; Kutas & Hillyard, 1984) target word with the same sentences containing either a low cloze probability target word (weakly constraining sentence context) or a semantic violation. They showed a gamma power increase only in the high cloze condition. The findings discussed all show that gamma power increases whenever semantic unification is successful.

Some recent studies have begun to investigate how power in the beta and gamma frequency ranges evolves over the course of an unfolding sentence, and how this might change when sentence processing is disrupted. Bastiaansen et al. (2010) compared syntactically legal sentences to sentences containing a syntactic violation (word category violation), and to randomised lists of the words contained in the legal sentences (little or no syntactic structure). They showed a linear increase in beta power across the sentence for the syntactically legal sentences and for sentences containing a syntactic violation, but only up to the point of the violation, at which time beta power started to return to baseline levels. There was no increase in beta power for the randomised word list condition containing no syntactic structure. Similarly, Bastiaansen and Hagoort (2015) compared syntactically and semantically legal sentences to randomised word lists (global syntactic violation) and to syntactic prose (syntactically legal sentences that are semantically uninterpretable because all content words are replaced by other unrelated content words; global semantic violation) within the same set of participants. They replicated the findings from Bastiaansen et al. (2010) for beta power (although this time there was a linear decrease across the sentence for the randomised word list condition and no linear trend for the syntactically legal sentences), and showed that gamma power was higher for semantically legal sentences compared to syntactic prose, but that there was no linear trend across the sentence in the case of the relationship between gamma power and semantic processing. Based on these findings, Bastiaansen and Hagoort (2015) proposed the “frequency-based segregation of

syntactic and semantic unification” hypothesis, suggesting that synchrony in the beta and gamma frequency ranges might constitute separate channels for the simultaneous processing of syntactic and semantic information during language comprehension.

On the other hand, it is not clear that the “frequency-based segregation of syntactic and semantic unification” hypothesis holds for all available data. In a similar paradigm, Wang, Jensen, et al. (2012) observed a beta (but not gamma) power difference when comparing target words in semantically anomalous sentences to those in semantically acceptable sentences. There are a number of other examples (see Lewis et al., 2015 for discussion) of cases where a strict beta-syntax and gamma-semantics link does not appear to hold. While it therefore appears clear that there is a link between both beta and gamma oscillatory neural activity and sentence-level language comprehension, whether those links are exclusive to syntactic and semantic aspects of sentence processing respectively remains less clear.

Another electrophysiological signature of brain activity that is sensitive to semantic processing is the N400 ERP component (see Kutas & Federmeier, 2011 for a recent review). The N400 is characterised by a negative-going deflection in the ERP waveform, typically peaking around 400 ms after the onset of a target word. An N400 can be observed in response to all content words in a sentence (along with other potentially meaningful stimuli), and the amplitude of the deflection is sensitive to a number of factors, most important of which for the present study is how easily a target word can be integrated into some preceding sentence context (cloze probability; Kutas & Federmeier,

2011). When comparing target words in a sentence with high cloze probability (good semantic fit; e.g. *The peanut was salted*) to words with low cloze probability (e.g. *The peanut was small*), or to semantically incongruous words (e.g. *The peanut was in love*), the amplitude of the N400 is reduced and the difference between conditions is termed an N400 effect, exhibiting a characteristic centro-parietal scalp distribution.

More recently it has been shown that discourse-level information can have an influence on local semantic processing within a sentence (see van Berkum, 2012 for review). For instance, Nieuwland and Van Berkum (2006) showed that by inserting a sentence containing a semantic animacy violation into a discourse context that changes reader’s/listener’s expectations about the animacy of the discourse referent (e.g. in the sentence from the previous paragraph *The peanut* is described in the preceding discourse as having animate characteristics), the direction of the N400 effect can be reversed (the N400 is now more negative when *salted* is the target word compared to *in love*). This is evidence that discourse-level information can have an effect on electrophysiological signatures related to the processing of semantics within a sentence.

Something that is not yet clear is exactly which discourse-level factors (only animacy or possibly other factors like discourse coherence or anaphora) can influence ERP signatures related to sentence-level processing, and under which circumstances. Another outstanding question concerns whether or not discourse-level factors can influence the oscillatory signatures that have been found for sentence-level semantic and syntactic unification. In the present study, we aimed to address some of these questions by revisiting an existing data set where participants read semantically coherent (COH) and incoherent (INCOH) short stories while their EEG was measured (Lewis, 2012). COH stories consisted of four sentences that fit together to describe a situation or event, and INCOH stories consisted of four unrelated sentences (see Table 1 for example stories).

Our original approach to the analysis, investigating the evolution of power over the course of the entire story, did not yield any statistically significant results (Lewis, 2012; see also Lewis et al., 2015 for discussion). We have argued that one reason for this may be the relatively poor SNR in the data due to the unusually long trials (23,200 ms for the word by word presentation of the whole story). In the present study, we improve the SNR by taking the average over all words within each sentence and each condition, thus retaining temporal information in the form of the sentence number (first, second, third, or fourth sentence of the story), but averaging out as much noise as possible. This approach has

Table 1. Example materials and their English translation (in italics).

Condition	Example materials
COH	Charles verliet zijn vaderland Senegal om in Europa te werken. <i>Charles left his home country Senegal to work in Europe.</i> Met een levensgevaarlijk klein bootje werd hij naar Tenerife gesmokkeld. <i>With a dangerously small boat he was smuggled to Tenerife.</i> Hij moest daar hard werken voor een klein beetje geld. <i>There he had to work hard for very little money.</i> Zijn familie had het geld dat hij stuurde hard nodig. <i>His family desperately needed the money that he was sending.</i>
INCOH	Charles verliet zijn vaderland Senegal om in Europa te werken. <i>Charles left his home country Senegal to work in Europe.</i> Een avond hadden ze een taart achtergelaten in de keuken. <i>One evening they left a hot pie in the kitchen.</i> Toevallig kwam een agent de hoek om die hen arresteerde. <i>Coincidentally a cop came around the corner that arrested them.</i> Maar na een jaar moest hij al naar de sloop. <i>But after just a year it was ready for the dump.</i>

Notes: COH: semantically coherent condition; INCOH: semantically (discourse-level) incoherent condition; *italics*: English translation.

the added benefit that it addresses the within-sentence temporal variability of a potential effect of our experimental manipulation. Although we can be certain that discourse coherence breaks down at sentences 2–4 for the INCOH condition (see *Methods* section, rating task), it turned out to be impossible to tightly control the exact point within each sentence at which this occurred (e.g. coherence might break down at word number 5 of the fourth sentence for one stimulus item but at word number 7 of the fourth sentence for another stimulus item, and the point of coherence break down for a given stimulus will presumably even vary between participants). By averaging over all words within each sentence we effectively remove this issue, at the cost of a potential loss of sensitivity due to the inclusion of words where an effect of our manipulation does not occur. This likely makes our statistical analyses particularly insensitive to the detection of potential interaction effects.

We performed a time–frequency (TF) analysis of power changes relative to a pre-story baseline period in partially overlapping low (2–30 Hz) and high (28–100 Hz) frequency ranges. This allows us to test whether discourse-level semantic coherence has any effect on local sentence-level processing, reflected in differences in beta and/or gamma power. We also performed an ERP analysis as we suspected that the N400 might be sensitive to our semantic coherence manipulation. The ERP analysis was also performed with the data averaged over all words within each sentence and each condition.

We had two main hypotheses for this experiment. First, we hypothesised that the N400 ERP component should be sensitive to our semantic manipulation, and thus should be larger for the INCOH than the COH condition at sentences 2–4. This will serve as an indication of whether or not our manipulation of discourse semantics has any effect on online semantic processing. Second, we expect to find higher gamma and/or beta power in the COH than in the INCOH condition for sentences 2–4. Gamma power modulations related to our experimental manipulation would clearly reflect disrupted local semantic processing. Modulations of beta power on the other hand are less clear, because as we have outlined above, beta has been observed for manipulations of both syntactic and semantic processing.

Methods

Participants

Thirty native speakers of Dutch took part in the experiment, 20 of whom were included in the final analysis

(7 males, 13 females; aged 18–27). Participants provided informed consent and were paid or equivalently rewarded with course credits for their participation. All participants reported normal or corrected-to-normal vision, and were right handed. None of the participants reported any neurological impairment, nor had they participated in any of the previous experiments involving the construction of the stimulus materials.

Five participants were excluded from the final analysis due to recording problems. Five other participants were excluded due to poor data quality (more than 37.5% of trials rejected in either condition). The relatively high number of excluded participants is primarily due to the fact that the experimental trials had a considerably longer duration than in experiments on isolated sentences, thus leading to more eye movements and other movement artefacts.

Stimulus materials

All stimuli consisted of Dutch short stories, each comprising four syntactically and semantically acceptable sentences. Every sentence contained exactly 10 words (Table 1). Conditions differed in terms of whether the sentences comprising the stories formed a coherent discourse (COH), or were unrelated to one another (INCOH). Two additional conditions where the discourse became incoherent starting at the third and fourth sentences respectively were included as fillers.

Eighty coherent and 80 incoherent stories (specifications just described) were taken from Hoffmann (2011). For the INCOH condition, the second, third, and fourth sentences in the COH stories were randomly exchanged across items (i.e. sentence 2 from the first story was exchanged with sentence 2 from one of the other 79 COH stories; sentence 3 from the first story was exchanged with sentence 3 from one of the other 79 COH stories; etc.). Latent Semantic Analysis (LSA) scores (Landauer, Foltz, & Laham, 1998) were calculated to confirm that this led to low semantic coherence between the sentences in this condition (Hoffmann, 2011). These 80 COH and 80 INCOH stories were counter-balanced across participants using two experimental lists, each comprising 40 COH and 40 INCOH stories. This meant that no participant read the same sentence more than once in the experiment, while all sentences appeared an equal number of times at the same sentence position in both conditions. Importantly, this shuffling procedure was applied at the level of sentences, so that word position within a sentence and overall sentence structure was preserved across participants. The only factor that was manipulated for each of sentences 2–4 was whether or not that sentence fit coherently

with the preceding sentence (or sentences). For the fillers, the third and fourth (10 stories), and fourth (10 stories) sentences respectively of an additional 20 coherent stories not used in the main comparisons were randomly exchanged to induce a breakdown in semantic coherence starting at sentence 3 or at sentence 4.

A rating task was performed with a group of participants who did not take part in the EEG experiment (Hoffmann, 2011). Their task was to rate how well each sentence comprising the stories fit with the previously presented sentence on a 7-point Likert scale, with 1 being “not at all”, and 7 being “very well”. The results¹ are shown in Table 2 and indicate that for the INCOH and two filler conditions participants were already aware of the semantic coherence breakdown at the end of the second, and third or fourth sentences of the stories respectively.

Experimental design and procedure

Participants were tested in a dimly lit, sound-attenuating, and electrically shielded booth. They were seated comfortably in front of an LCD computer monitor (Samsung SyncMaster 940 eW), with a viewing distance of between 70 and 80 cm. Letters were presented in white on a black background using a 20-point sized Consolas font type. All words subtended a visual angle of 3.13° vertically.

Sentences were presented word by word in the centre of the screen. For each sentence, the first letter of the first word was capitalised, and the final word was presented with a period. A single trial consisted of an entire story containing four sentences, a movement cue, and a fixation cross. Words were presented for 300 ms, followed by a 200 ms blank screen between words. Each trial began with the presentation in the centre of the screen of three asterisks separated by two spaces for 4000 ms, indicating that participants could move their eyes and blink. This was immediately followed by a fixation cross presented in the centre of the screen for 3000 ms, indicating that eye movements and blinking should be avoided, and that the story was about to

Table 2. Results from the rating task.

Condition	Sentence 2		Sentence 3		Sentence 4	
	Mean	SD	Mean	SD	Mean	SD
COH	6.6	0.4	6.6	0.3	6.7	0.3
INCOH	2.3	0.5	1.7	0.5	1.4	0.3
INCOH 3	6.4	0.6	1.6	0.6	1.3	0.4
INCOH 4	6.3	0.5	6.3	0.6	1.6	0.5

Notes: COH: semantically coherent condition; INCOH: semantically (discourse-level) incoherent condition; INCOH 3 and INCOH 4: fillers; rating of 7: fits perfectly with previous sentence; rating of 1: does not fit with previous sentence at all.

begin. The first word of the first sentence immediately followed the fixation cross. Each sentence lasted 5000 ms and was followed by an 800 ms inter-sentence blank screen before the onset of the first word of the next sentence. For the last sentence of each story, the inter-sentence interval was immediately followed by a new trial. A single trial lasted 30,200 ms (including fixation and blinking periods).

Participants were instructed to read all stories attentively for comprehension, and to continue reading regardless of whether or not the story made sense to them. They read a total of 100 stories (40 COH, 40 INCOH, and 20 fillers), presented in 20 blocks of 5 stories each. For 50% of the participants, stories in an experimental list were presented in reverse order to control for potential order effects. Ten training stories were presented to participants before the experiment.

EEG recordings

Participants were fitted with a 64-electrode actiCap with electrodes positioned according to the standard 10/20 system. EEG signals were recorded using 60 Ag/AgCl active sensors mounted in the cap and referred to the right mastoid. An additional electrode was placed on participants' left mastoid for re-referencing offline, and a ground electrode was placed on the centre of the forehead. An additional electrode was placed on the sub-orbital ridge of participants' left eye for recording eye-blinks.

Electrode impedance was kept below 10 k Ω . EEG and electrooculogram (EOG) recordings were amplified using BrainAmp DC amplifiers (Brain Products GmbH, Gilching, Germany) with a band-pass filter of 0.053–249 Hz, digitised online with a sampling frequency of 1000 Hz and stored for offline analysis.

Data pre-processing

EEG data were analysed using the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) running in a MatLab environment (R2012a; Mathworks, Inc.). For each participant, scalp electrodes were re-referenced to the average of electrodes placed on the left and right mastoid (linked-mastoid reference). A band-stop filter was applied at 50, 100, and 150 Hz in order to minimise the effects of power line interference (50 Hz) and data were segmented from –2500 to 24,000 ms relative to the onset of each story.

Next the data were decomposed into independent components (independent components analysis [ICA]) using the “runica” implementation in FieldTrip with default settings, resulting in 60 components time

courses. Components which captured eye-blinks and eye movements were removed, and the remaining components were recombined (Jung et al., 2000; Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997). Between 0 and 2 components were removed per participant.

Trials still containing muscle artefacts were rejected by visual inspection of the data, band-pass filtered between 110 and 140 Hz (this frequency range is where the majority of the energy associated with muscle activity is concentrated). Any linear trends in the data were removed and a baseline correction was applied to every trial using an interval from -2500 to 0 ms relative to story onset. Trials with amplitude higher than $75 \mu\text{V}$ or lower than $-75 \mu\text{V}$ were excluded from further analysis. There was no significant difference between the number of trials in the two conditions (COH: $M = 29.4$, $SD = 2.72$; INCOH: $M = 29.05$, $SD = 3.24$; $p = .66$). Finally, data were segmented from -1000 to 1000 ms relative to the onset of every word of each sentence comprising the stories for the COH and INCOH conditions combined.

ERP analysis

Data for each word were high-pass filtered above 0.1 Hz and low-pass filtered below 30 Hz using a windowed sinc finite-impulse response filter with FieldTrip default settings. Next, data were segmented into COH and INCOH conditions, and within each condition into sentences 1–4 respectively. A condition-specific baseline correction using a period from -200 to 0 ms relative to word onset was applied to the data for each individual word. Finally, data were averaged within each of the four sentences for the two conditions separately from -100 to 500 ms relative to word onset to obtain participant-specific ERP waveforms.

TF and inter-trial coherence analyses

A multitaper approach (Mitra & Pesaran, 1999) was used to compute TF representations for the single trial data of each participant. TF representations were calculated in two partially overlapping frequency ranges (Womelsdorf, Fries, Mitra, & Desimone, 2006) because of the trade-off between time and frequency resolution that results from this multitaper approach.

In a high-frequency range (28 – 100 Hz), 250 ms time-smoothing and 8 Hz frequency-smoothing windows applying a Slepian taper sequence were used to calculate power changes in frequency steps of 4 Hz and time steps of 10 ms. Each time point in the resultant Time-Frequency Representations is thus a weighted average of the time points ranging from 125 ms before to

125 ms after this time point. In a low-frequency range (2 – 30 Hz), 400 ms time-smoothing and 2.5 Hz frequency-smoothing windows using a Hanning taper were applied in frequency steps of 1 Hz and time steps of 10 ms.

Single-trial Fourier spectra were averaged for each participant from 0 to 500 ms relative to the onset of each word of the stories. Data epoch length is limited by the stimulus onset asynchrony (SOA) of 500 ms in order to avoid averaging overlapping data segments. This resulted in a TF representation of power for each participant averaged over all words comprising the stories, irrespective of condition or sentence number. These participant averages were then expressed as a relative change (in dB) from a baseline period between 750 and 250 ms prior to story onset (fixation). The average TF representation of power over all participants and scalp electrodes was then calculated for visual inspection.

For the low-frequency range, we also computed the inter-trial coherence (ITC; Tallon-Baudry, Bertrand, Delpeuch, & Pernier, 1996) for each participant from 0 to 500 ms relative to the onset of each word of the stories, by first normalising the Fourier spectrum of each trial by its amplitude and then averaging the result across all trials for that participant. This provides a frequency-resolved measure of the degree of trial-to-trial phase consistency over time (Makeig et al., 2004). We used this to distinguish evoked activity (strongly phase-locked and likely related to the ERP results) from induced activity (time- but not phase-locked) in subsequent TF analyses. Resultant participant-specific ITC values were then averaged over all participants and scalp electrodes for visual inspection.

Next, we selected TF ranges of interest for the low-frequency range based on previous literature and on visual inspection of the TF and ITC data averaged over all words, all participants, and all scalp electrodes (Figure 2(a)). In this way we were able to select regions of interest without statistical comparison between conditions at this stage (there is no condition-specific information in this representation of the data), thus avoiding double-dipping later in the statistical analysis. Our criteria were (1) a visible increase or decrease in power in the TF representation relative to baseline (to establish the presence of a power modulation compared to baseline); (2) only weak (less than 0.15) or no phase-locking visible for the corresponding TF range in the ITC values (to exclude the potential contribution of ERP components to the TF representations of the induced oscillatory activity of interest); and (3) good correspondence with previous results in terms of frequency range selected (to ensure any potential effects make sense in light of previous

literature). This resulted in the selection of alpha/theta (6–12 Hz; 350–500 ms relative to word onset), alpha/beta (9–17 Hz; 0–140 ms relative to word onset), early beta (21–28 Hz; 40–160 ms relative to word onset), and late beta (13–21 Hz; 260–480 ms relative to word onset) TF ranges of interest for further analysis (see black boxes in Figure 2(a)). In addition, we selected a TF range of interest for the large theta power increase (4–7 Hz; 150–300 ms) despite the ITC data clearly indicating that this is phase-locked activity, in order to investigate whether or not this is related to potential ERP findings (for discussion see Bastiaansen, Mazaheri, & Jensen, 2012).

For the high-frequency range, we performed statistical analyses on the entire range (28–100 Hz), as well as for the mean power in a low gamma frequency range (35–55 Hz) based on the majority of findings relating gamma power to semantic processing (Lewis et al., 2015).

Single-trial Fourier spectra per participant were then segmented into COH and INCOH conditions, and within each condition into sentences 1–4, from 0 to 500 ms relative to word onset. Fourier spectra were averaged, resulting in participant-specific averages for sentences 1–4 for the COH and INCOH conditions respectively. These participant averages were then expressed as a relative change (in dB) from the baseline period between 750 and 250 ms prior to story onset (fixation). This provides us with a measure of the average relative power change from baseline for all words within each sentence of the stories for the COH and INCOH conditions separately. By averaging over all words in a sentence, we improve the SNR and at the same time take into account temporal variability in the potential effect of our coherence manipulation on theta, alpha, and beta power (e.g. appearing at word 7 in one trial and word 8 in another trial).

Statistical analyses

The statistical significance of all comparisons was evaluated using a cluster-based random permutation approach (Maris & Oostenveld, 2007). We used this approach because of its natural handling of the multiple comparisons problem.

Cluster-based random permutation statistics control the family-wise error rate by making use of the spatial, spectral, and temporal autocorrelation in EEG data. In short, a dependent samples *T*-test is performed for every data point (electrode–frequency–time point for TF or electrode–time point for ERP data) giving uncorrected *p*-values. A pre-set significance level is chosen (here 5% single-tailed for the N400 ERP analysis; 5%

two-tailed, for all other comparisons) and any data points not exceeding this level are discarded (set to zero). Clusters are calculated from the remaining data points based on their adjacency in space (adjacent electrodes), time, and frequency. Cluster-level statistics are then calculated by summing the values of the *T*-statistics for all data points in each cluster. A permutation distribution is created by randomly assigning participant averages to one of the two conditions 3000 times, and each time calculating cluster-level statistics as just described. The highest cluster-level statistic from each randomisation is entered into the permutation distribution and the cluster-level statistics calculated for the measured data are compared against this distribution. Clusters falling in the highest or lowest 2.5th percentile of the estimated distribution were considered significant (lowest 5th percentile for the N400 ERP analysis).

We compared the COH and INCOH conditions separately for each sentence comprising the stories. For the ERP data we hypothesised that our manipulation of discourse semantics should result in a more negative N400 peak in the INCOH than in the COH condition for sentences 2–4. Our statistical comparison was therefore based on the mean ERP amplitude in a time window (300 to 500 ms relative to word onset) typically capturing N400 effects (see e.g. Kutas & Federmeier, 2011). To test whether there were any earlier ERP effects we tested the time window between 0 and 300 ms relative to word onset, now forming clusters in time as well as space. For the TF data we compared mean power values in the selected theta (4–7 Hz; 150–300 ms relative to word onset), alpha/theta (6–12 Hz; 350–500 ms relative to word onset), alpha/beta (9–17 Hz; 0–140 ms relative to word onset), early beta (21–28 Hz; 40–160 ms relative to word onset), and late beta (13–21 Hz; 260–480 ms relative to word onset) TF ranges, forming clusters only in space. We also compared COH and INCOH conditions for the entire gamma frequency range (28–100 Hz), clustering in space, frequency, and time, as well as in a low gamma frequency range of interest (35–55 Hz), clustering in time and space.

For observed statistically significant differences between COH and INCOH conditions in both the TF and ERP analyses we tested for an interaction between condition (COH/INCOH) and sentence position (Sentence1/Sentence2/Sentence3/Sentence4) by extracting mean power or amplitude values respectively in the TF region or time window of interest, averaged over all electrodes identified based on the output of the cluster-based statistics from the sentence exhibiting the largest effect. These values were entered into a repeated measures analysis of variance with condition and sentence position as factors. There were no cases where

sphericity was violated according to Mauchley's test. We only interpreted interaction effects, but not main effects, in order to avoid double-dipping. Interactions were not broken down further since we already have pairwise comparisons from the cluster-based statistical output.

Results

Statistical comparisons were made between COH and INCOH conditions separately for each of the four sentences comprising the stories. Interactions between sentence position and condition were also tested. We expected our statistical analyses to be relatively insensitive to potential interaction effects, due to the poor SNR and averaging over all words within a sentence (as already discussed, effects of manipulating discourse-level semantic coherence are not likely to be present at all words of the sentences). We therefore still describe differences between conditions at each sentence position, but are careful not to make inferential claims about how this differs from sentence to sentence in cases where the interaction is not significant.

ERP results

ERP effects were quantified by differences in mean amplitude in the N400 time window (300 to 500 ms relative to word onset), or by temporally and spatially contiguous time points identified by the cluster-based permutation approach in an earlier time window (0 to 300 ms relative to word onset). [Figure 1](#) shows the ERP waveforms for the COH and INCOH conditions for each of the four sentences comprising the stories (left column), along with the scalp distribution of the difference between conditions (INCOH minus COH) for the early and N400 time windows (middle and right columns respectively).

N400 time window

In the N400 time window we observed a statistically significant negative difference between the INCOH and COH conditions at the third ($p = .03$), and fourth ($p < .001$) sentences, with this difference exhibiting a trend towards significance at the second sentence ($p = .05$). The interaction between condition and sentence position was statistically significant ($p = .016$). This, in combination with the cluster-based pairwise comparisons, suggests that sentences 2–4 exhibit differences between COH and INCOH conditions, while the first sentence does not. The INCOH condition exhibits a larger negative-going deflection compared to the COH condition between about 350 and 450 ms after word onset, with a centro-parietal scalp distribution for the

difference ([Figure 1](#)). Based on the timing and scalp distribution we identify this as an N400 effect.

Early time window

In the early time window, we observed a significant positive difference between the INCOH and COH conditions at the third ($p = .02$) and fourth ($p = .02$) sentences. The interaction between condition and sentence position was not statistically significant ($p = .75$). Based on these results we cannot make any inferential claims about whether the difference between COH and INCOH conditions differs across sentence positions. Any discussion of such effects for the ERP data in the early time window is thus purely descriptive in nature. The COH condition shows a larger negative-going deflection compared to the INCOH condition over a wide range of electrodes ([Figure 1](#)). This effect appears to begin around 80 ms relative to word onset and lasts until about 200 ms. Closer examination reveals that this difference is already present at the second sentence, but that there it is smaller and less widely distributed across the scalp. As a result, it is not significant there ($p = .22$). Based on the timing and scalp distribution of this difference we argue that it is likely a visual N1 effect.

TF results

For the low-frequency range, we selected five TF ranges of interest for statistical comparison, a theta (4–7 Hz; 150–300 ms relative to word onset), an alpha/theta (6–12 Hz; 350–500 ms relative to word onset), an alpha/beta (9–17 Hz; 0–140 ms relative to word onset), an early beta (21–28 Hz, 40–160 ms relative to word onset), and a late beta (13–21 Hz; 260–480 ms relative to word onset) range. [Figure 2\(a\)](#) shows the TF representation of power (top) and corresponding ITC values (bottom) averaged over all words in the stories irrespective of condition or sentence. TF ranges of interest are marked by black boxes.

There were no statistically significant differences between the COH and INCOH conditions for any of the four sentences in the theta, the alpha/theta, the alpha/beta, or the early beta TF ranges of interest. In the late beta TF range there was a significant difference between COH and INCOH conditions at the fourth sentence of the stories ($p = .02$). The interaction between condition and sentence position was not statistically significant ($p = .86$). Based on these results we cannot make any inferential claims about whether the difference between COH and INCOH conditions differs across sentence positions. Any discussion of such effects for the TF data in the late beta frequency range is thus purely descriptive in nature. [Figure 2\(b\)](#) shows the scalp distribution of the difference between

conditions (COH minus INCOH) for each of the four sentences in the late beta TF range, along with bar plots of the power decrease for each condition at sentences 1–4, averaged over all electrodes contributing to the statistically significant difference at the fourth sentence. Sentence 4 clearly exhibits the strongest and most widespread difference, with maxima over left fronto-central, right frontal, and right temporal electrodes. Sentence 3 also exhibits a difference, but it only shows a trend towards statistical significance ($p = .06$), with maxima over left fronto-central and right temporo-parietal electrodes. The effect is driven by a decrease in beta power relative to baseline in both conditions (Figure 2(a) and bar plots in Figure 2(b)), which is stronger in the INCOH than in the COH condition.

For the high-frequency range, we selected the entire frequency range (28–100 Hz) as well as a low gamma frequency range of interest (35–55 Hz) for statistical comparison. There were no statistically significant differences between the COH and INCOH conditions for either of these ranges in any sentence position. This suggests that local sentence-level gamma power was not sensitive to our manipulation of discourse-level semantic coherence.

Correlation analysis

In order to test whether there is any direct relationship between our beta oscillatory effects and our N400 ERP findings, we performed a Pearson correlation between N400 amplitude differences and beta power differences between the COH and INCOH conditions for sentences 2–4. For every participant, and for each sentence separately, we extracted mean power difference values from the late beta TF range of interest, averaged over all electrodes contributing to the statistically significant difference at the fourth sentence. We also extracted mean amplitude differences per participant for each sentence, averaged over the N400 time interval and over all electrodes contributing to the statistically significant difference at the fourth sentence. A Pearson correlation was then performed across participants between these two values. This correlation analysis did not produce any statistically significant correlations between beta power and N400 amplitude differences. We can thus conclude that there is no direct relationship between our beta oscillatory and N400 effects related to our experimental manipulation of discourse-level semantic coherence.

Discussion

The frequency-based segregation of syntactic and semantic unification hypothesis (Bastiaansen & Hagoort, 2015) claims that there is a close relationship

between syntactic processing and oscillatory neural dynamics in the beta frequency range, and between semantic processing and oscillatory activity in the gamma frequency range. On the other hand, modulations of beta power have also been observed for manipulations of semantic processing (e.g. Wang, Jensen, et al., 2012). Furthermore, discourse-level information can influence electrophysiological brain signatures related to local sentence-level semantic processing (i.e. the N400 ERP component; Nieuwland & Berkum, 2006). We tested whether a different discourse-level factor, semantic coherence, can influence local sentence-level processing as indexed by modulations of the N400, as well as beta and gamma oscillatory power. Participants read groups of four sentences that either formed coherent stories (COH) or were semantically unrelated (INCOH), as indicated by both LSA scores (Hoffmann, 2011) and a rating task (Table 2).

The ERP analysis produced an N400 effect at sentences 2–4, with more negative-going waveforms in the INCOH compared to the COH condition, and a visual N1 effect that was only significant at sentences 3 and 4, with a more negative peak for the COH compared to the INCOH condition (Figure 1). The TF analysis of power produced a single result in the beta frequency range (13–21 Hz), with higher beta power in the COH compared to the INCOH condition at sentence 4 between 260 and 480 ms relative to word onset, exhibiting a frontal, fronto-central, central, and centro-parietal scalp distribution. At sentence 3 a trend towards significance was present, exhibiting a similar scalp distribution. There were no gamma power effects related to our manipulation of sentence-level semantic coherence. Importantly, we observed an interaction between sentence position and condition only for the N400 effect, and thus any discussion of differences between the COH and INCOH conditions differing across sentence position for the N1 ERP and for the beta TF findings is purely descriptive in nature.

Discourse-level semantic coherence influences the N400

The main purpose of the ERP analysis was to confirm that our discourse-level semantic coherence manipulation had an effect on online sentence-level semantic processing. We deliberately avoid entering discussions about the nature of the processing giving rise to N400 effects (e.g. lexical retrieval difficulties, disruption of semantic integration, or lower predictability; see Kutas & Federmeier, 2011), as we do not think our experimental manipulation allows us to add anything to this debate.

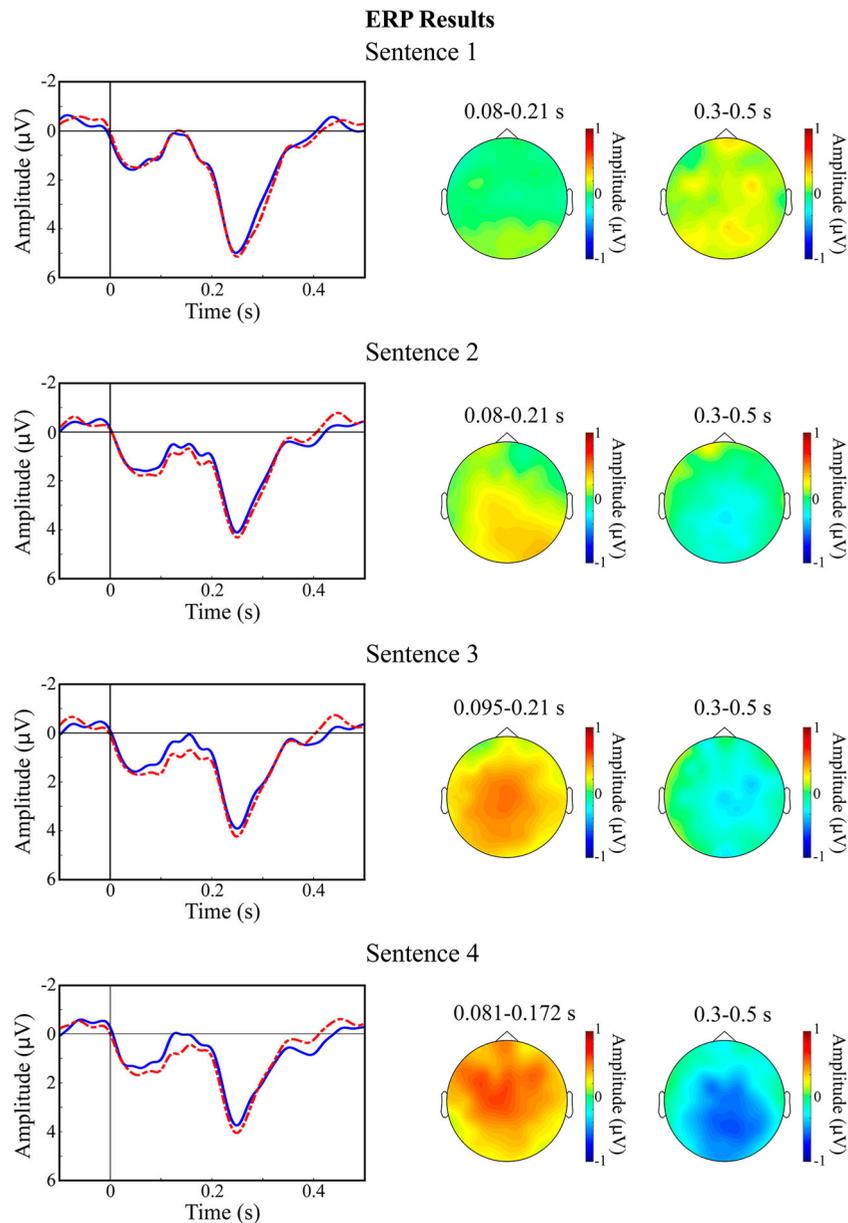


Figure 1. Results of the ERP analysis. The results for the average over all words in each sentence are presented separately. The left column shows ERP time courses for a representative electrode CPz (blue solid = COH; red dashed = INCOH). The middle column shows the scalp distribution of the difference between conditions (INCOH minus COH) averaged over a time window corresponding to the significant effects (sentences 3 and 4) found in the early time window analysis. The third column shows the scalp distribution of the difference between conditions (INCOH minus COH), now averaged over the N400 time window (300–500 ms relative to word onset), exhibiting a statistically significant effect at sentences 3 and 4, and a trend at sentence 2.

As hypothesised, the N400 was sensitive to our discourse-level semantic manipulation as soon as the stories became incoherent (sentences 2–4). It is clear from the scalp distribution in Figure 1 (right-most column) that the effect becomes larger and more widespread across the scalp (this does not necessarily indicate that more brain regions become involved, but can also result from increased activity at the same underlying sources) as we move from sentence 2–4. The N400 is classically related to semantic processing at the level of

single words, as well as the sentence level (and also certain non-linguistic stimuli; Kutas & Federmeier, 2011), and has also been shown to be sensitive to discourse-level semantic information in the form of the animacy assigned to a particular referent based on the preceding discourse (Nieuwland & van Berkum, 2006). Here, we show that a different discourse-level factor, semantic coherence between sentences comprising a short story, can also have an influence on the N400. We argue that this is related to local sentence-level

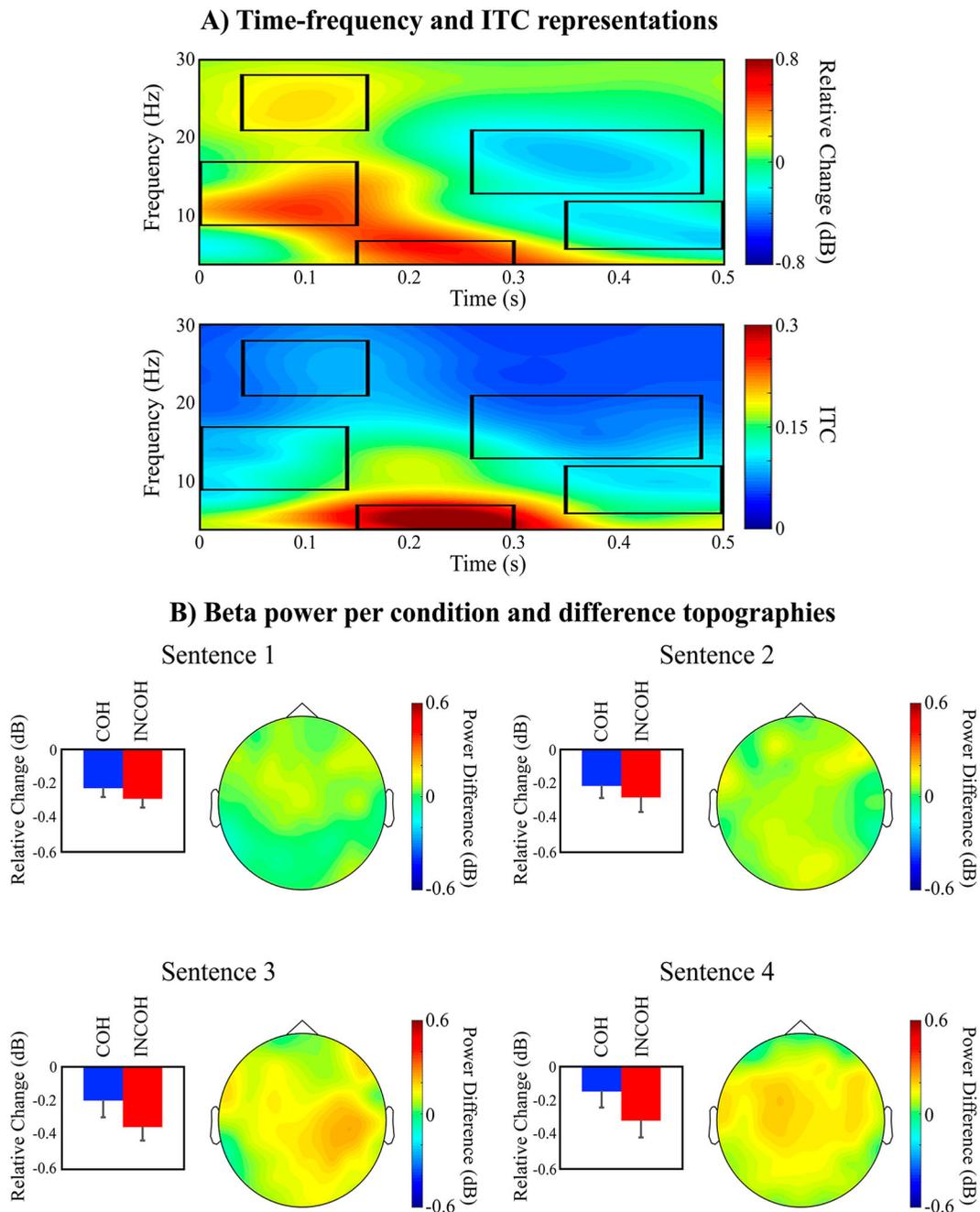


Figure 2. Results of the TF analysis of power. (a) TF (top) and ITC (bottom) representations for the average over all participants, all scalp electrodes, and all words in all sentences in both the COH and INCOH conditions. Black boxes indicate the TF ranges of interest selected for statistical testing. (b) Bar plots of mean power in the beta TF range of interest (13–21 Hz; 260–480 ms relative to word onset) for each condition averaged over electrodes contributing to the statistically significant difference at sentence 4, as well as scalp distributions of the difference between conditions (COH minus INCOH) averaged over all words within each of the four sentences for the beta TF range of interest. This difference is statistically significant at sentence 4 and shows a trend towards significance at sentence 3. Error bars on the bar plots indicate standard error of the mean.

semantic processing. Bearing in mind that we average over all words within each sentence for the two conditions separately, it is likely that our manipulation of semantic coherence has an effect on the N400 at one, or more likely at a few words within each sentence (probably a variable number of words in each sentence across

items, and perhaps even participants), and that this is what drives the effect when averaging over all words for sentences 2–4. This would suggest that manipulating semantic coherence at the discourse level disrupts local sentence-level semantic unification, possibly related to difficulties with processing anaphoric relations from

sentence to sentence (van Berkum, Koornneef, Otten, & Nieuwland, 2007) and/or with thematic role assignment (Paczynski & Kuperberg, 2011). A question that should be addressed in future work is which of these aspects of sentence-level processing is affected by manipulating discourse-level semantic coherence. For instance, more fine-grained stimuli could be used to specifically target the breakdown of anaphoric relations between sentences (e.g. Koornneef & Sanders, 2013).

St. George, Mannes, and Hoffinan (1994) report a similar finding for ambiguous paragraphs preceded by a disambiguating title compared to the same paragraphs without a title. The presence of a title causes the sentences comprising the paragraphs to be more coherent, and disambiguates the paragraph for the reader. They report an N400 effect, with a more negative-going N400 component for the condition without a title (ambiguously related sentences) compared to the condition with a title (coherent sentences), exactly in line with our N400 effect. Importantly, they also report an enhanced P1-N1 ERP component for their condition with a title. They interpret this finding in terms of a better ability to allocate attention to the condition where sentences are more coherent due to the presence of a title. This directly corroborates our visual N1 effect at sentences 3 and 4 (Figure 1). We observed a larger N1 in the COH compared to the INCOH condition between about 80 and 200 ms relative to word onset. Since this ERP component has been linked to the allocation of visual attention (e.g. Vogel & Luck, 2000), we conclude that our participants allocate more attention to the words in each sentence in the COH condition compared to the INCOH condition for sentences 3 and 4. This is likely because that information is more relevant in the COH condition, when a detailed situation model has to be constructed, whereas in the INCOH condition sentences can simply be read without any further processing related to construction of a detailed situation model. This becomes clearer the later in the stories a sentence appears, which might be the reason the N1 difference appears to become larger at later sentences (Figure 1).

Higher beta power for coherent short stories

Our manipulation of discourse-level semantic coherence had an effect on beta power at the third and fourth sentences of the stories (Figure 2(b)), but only reached statistical significance at the fourth sentence. The effect at the fourth sentence is widely distributed across the scalp, with maxima over left fronto-central and right temporal electrodes. At the third sentence the difference shows maxima over left fronto-central and right temporo-parietal electrodes. These positive differences

are the result of a larger decrease in power relative to baseline for the INCOH compared to the COH condition at sentences 3 and 4. In the introduction, we linked oscillatory power in the beta frequency range to sentence-level syntactic processing, but pointed out that beta has also been modulated by manipulations of semantic processing. The results reported here might therefore be taken as evidence that discourse-level semantic information can influence either local sentence-level syntactic or semantic processing (or both). We have already argued that our N400 ERP effects are related to sentence-level semantic processing, which is influenced by the discourse-level semantic manipulation. A correlation analysis revealed no relationship between our beta power effects and the N400 ERP findings, suggesting that differences in beta power between COH and INCOH conditions is more likely related to an influence of discourse-level semantic coherence on local syntactic processing.

There are however alternative explanations for the difference in beta power that should also be considered. Weiss and Mueller (2012) suggest that besides binding (unification) during sentence processing, beta activity might play a role in the processing of action semantics, in memory-related processing, or in attention and the violation of expectations. Indeed motor-related beta activity has been shown when comprehending action-related language (e.g. van Elk, van Schie, Zwaan, & Bekkering, 2010; Moreno et al., 2015). We find it unlikely however that this could explain our beta findings, as we did not explicitly manipulate action-semantic content between our COH and INCOH conditions.

The argument for a relationship between beta oscillations and memory processes is based largely on biophysically realistic computational modelling, showing that increases in beta activity in local cortical circuits have exactly the characteristics that would be necessary to hold information online for extended periods of time (Kopell, Whittington, & Kramer, 2011). In contrast, a beta decrease in left frontal cortex has been linked to improved subsequent memory (Hanslmayr, Staudigl, & Fellner, 2012). It is possible that our COH condition placed higher demands on short-term memory than our INCOH condition because related sentences mean a situation model has to be maintained, whereas in the INCOH condition participants do not necessarily construct a situation model. On the other hand, it is also possible that our INCOH condition placed higher demands on short-term memory than our COH condition because unrelated sentences mean that if participants do attempt to construct a situation model after sentence 2, they may have to maintain more information in order to attempt to combine the sentences than in the COH

condition (where this is immediately possible). Each of these possibilities fits with one of the above ideas about the relationship between beta and memory processing. Other findings relating beta to short-term memory during language processing have all argued for a relationship between increased beta power and higher memory demands (Haarmann, Cameron, & Ruchkin, 2002; Meyer et al., 2013; Weiss et al., 2005), but these findings can be explained just as well by a link between beta activity and syntactic unification demands (see Lewis et al., 2015 for discussion). In fact, Meyer et al. (2013) explicitly argue that for their long- and short-distance subject–verb agreement dependencies *alpha* activity is an index of retention of information in short-term memory, while beta is related to syntactic integration, as it only appears at the end of the retention interval. Given that there is not yet consensus about the exact relationship between beta oscillations and short-term memory during language comprehension, and that all previous findings from language comprehension explicitly relating beta to short-term memory can be adequately explained in terms of differentially demanding syntactic processing, we think that our results are more straightforwardly explained by linking beta to local sentence-level syntactic processing.

Based on our findings, we cannot draw any strong conclusions about the exact mechanism by which discourse-level semantic information influences sentence-level syntactic processing. One very tentative proposal is that at points in a sentence when the language comprehension system becomes aware that the sentence-level meaning being constructed does not fit the wider discourse context, the system attempts a syntactic reanalysis in order to try to make the new information fit. This syntactic reanalysis would result in a disruption of local syntactic processing, leading to a decrease in beta power. If this occurred at multiple words within each sentence that would explain why the effect is present in our word averaged data. Our initial hypothesis was that gamma power would be modulated when sentence-level semantic processing was influenced by discourse-level semantic coherence, but we do not observe any gamma effects in the present study. As we have argued above, beta has also been linked to manipulations of sentence-level semantic processing, and this explanation of the findings cannot be fully ruled out.

The suggested link between beta activity on the one hand, and attention and expectancy violation on the other, is based on a broader proposal linking beta to the maintenance or change of the current cognitive processing set (Engel & Fries, 2010). These ideas have recently been made more explicit for the case of sentence-level language comprehension (Lewis et al.,

2015), where beta increases have been linked to the active maintenance of the brain network configuration responsible for the representation and construction of the current sentence-level meaning, while beta decreases have been linked to a change in the underlying network configuration when the system prepares for a new mode of processing. It is possible that for the INCOH condition at sentences 2–4 there are points in the sentence at which it becomes clear to the language comprehension system that the input does not fit coherently into the wider discourse or situation model. In such cases, the system might use these as cues, indicating that the current mode of processing (and thus the current sentence-level meaning) must change, and this would result in the observed lower beta power in the INCOH compared to the COH condition. It should be noted that this account of the findings in terms of maintenance/change of processing set is not incompatible with the account in terms of disrupted local syntactic or semantic processing, because as we have argued elsewhere (Lewis et al., 2015), if local syntactic processing is disrupted this should also act as a cue to the system that the current mode of processing has to change. In this experiment, we have not explicitly tried to disentangle these two accounts, and thus while we note that both accounts can adequately explain the beta findings, we prefer to interpret them in relation to the more specific link between beta and local sentence-level syntactic processing that is influenced by manipulating discourse-level semantic coherence.

One way we may have been able to shed more light on this issue is by looking into the P600 ERP component as a marker of syntactic integration (although it has also been shown that the P600 is not exclusively related to syntactic processing; e.g. van de Meerendonk, Kolk, Vissers, & Chwilla, 2010), but a limitation of our approach to the analysis (averaging over words within a sentence) is that only time points between 0 and 500 ms (the SOA) relative to word onset contain non-redundant information. This means that it would not have made sense to investigate the typical P600 time window.

A question that arises concerns why this beta effect is not (descriptively) found before sentence 3 and is only statistically significant at sentence 4, while the behavioural results clearly show that in the INCOH condition the stories already become incoherent at the second sentence. One possibility is that participants have to read to the end of the second sentence before they become aware of the breakdown in coherence in the INCOH condition (in the rating task responses are provided after each sentence). On the other hand, we do observe an N400 effect at the second sentence (Figure 1), indicating that participants' brains are already

sensitive to our discourse-level semantic manipulation during the second sentence. A more likely explanation is that the lack of coherence between sentences in the INCOH condition simply becomes clearer as one moves from reading sentence 2 to sentence 4. At the second sentence participants presumably notice the lack of coherence, but perhaps continue to try to combine the sentences anyway, whereas by the fourth sentence the lack of coherence is clear because sentences 2–4 were all incoherent with the first sentence and with each other. Although we did not test this explicitly, and the interaction between sentence position and condition was not statistically significant, visual inspection of the differences in beta power (Figure 2) along with the topographies associated with the N400 effects (Figure 1) lend support to this idea, as in both cases the difference seems to be larger and more widespread the later a sentence appears in the stories. As we have outlined earlier, our statistical analyses are likely relatively insensitive to interaction effects due to the poor SNR and due to averaging over all words within each sentence, many of which are likely to not exhibit effects related to our experimental manipulation.

Finally, our beta effect shows a difference over left fronto-central electrodes, and over right temporal and parietal electrodes. Although the spatial resolution with EEG is relatively poor, we may speculate that the left frontal difference is related to differential activation of left inferior frontal cortex, that may be disengaged in the INCOH condition when local sentence-level syntactic processing is disrupted (e.g. Hagoort, 2005, 2013; Meyer, Obleser, Kiebel, & Friederici, 2012; Tyler et al., 2011). The right hemisphere differences may be directly related to discourse-level semantic information, as it has been argued (Jung-Beeman, 2005) that the right hemisphere is involved in the recognition of more distant relations between, for example, discourse entities, and also to the activation of broader meaning (perhaps across sentences). In the INCOH condition this network of areas may be disengaged (hence the decrease in beta power) because no situation model can be constructed relating the sentences comprising the stories.

Limitations

This study has two major limitations that should be addressed in future research. First, we were forced to average over all words within each sentence for the two conditions in order to achieve a high enough SNR to observe effects in the data. While we have argued that this approach does retain temporal information in the form of the word averages for each of the four sentences comprising our stories, it does not take full

advantage of the EEG data related to the unfolding of each sentence of the stories. Indeed, our original intention was to investigate the evolution of power over the course of the entire discourse, and how that was modulated by discourse-level semantic coherence (Lewis, 2012). In addition, we rejected an unusually large number of participants (five) due to poor data quality. We have argued that the relatively poor SNR for our data is due to the unusually long trials, lasting for more than 20 s each. This makes it far less likely that participants are able to stay still during each trial, and when they strain to try to do so, likely results in an abundance of electromyographic (EMG) artefacts in the data. It was not possible to remove these using ICA, as they were present for many components, each time with different topographical configurations, often in combination with other non-artefactual activity. Similar future studies should attempt to reduce EMG artefacts by for instance presenting fewer trials for each recording block, and lengthening the inter-sentence interval in combination with explicitly encouraging participants to use the inter-sentence intervals to relax for a moment before keeping still again during the following sentence presentation. The poor SNR for our data may offer an explanation for the absence of any gamma effects related to our discourse-level semantic manipulation. EMG artefacts are especially likely to affect the ability to measure high-frequency oscillatory activity (e.g. Hipp & Siegel, 2013), and future studies investigating gamma should also consider analysing oscillatory power at the level of cortical sources, where spatial filters applied during source reconstruction (e.g. Gross et al., 2001) are likely to minimise the effects of EMG activity.

A second limitation is that we do not observe an interaction between sentence position and condition for our beta effect. This means that strictly speaking we cannot make inferential claims about whether the difference between COH and INCOH conditions is present for some sentences (sentences 3 and 4) but not for others (sentences 1 and 2). One potential reason that the interaction is not significant (we hypothesised no difference between COH and INCOH conditions at sentence 1, and a difference at sentences 2–4) is that we selected electrodes to test for the interaction based on those electrodes contributing to the significant difference between COH and INCOH conditions at sentence 4 from the output of the cluster-based permutation statistics. Figure 2(b) clearly indicates that the difference between conditions shifts around in space from sentence 3 to sentence 4, and this introduces additional variability into the data. In our opinion, it would not be appropriate to select

different electrodes for each sentence to test for an interaction between sentence position and condition, even though that might be most likely to produce the desired result. In addition, there is already a small difference between COH and INCOH conditions at the first sentence (Figure 2(b)), which while it cannot be an effect related to our experimental manipulation (there was no difference between COH and INCOH conditions at the first sentence), does go in the same direction as the effects at sentences 2–4, and so probably also contributes to the absence of an interaction between sentence position and condition.

Conclusion

This study shows that discourse-level semantic coherence has an effect on ERP and oscillatory responses related to local sentence-level language comprehension. The semantic processing difficulty that results when sentences are incoherent, leads to an enhanced N400 ERP component. In addition, oscillatory power in the beta frequency range is higher for coherent stories, potentially indicating that discourse-level semantic coherence also affects local syntactic processing. More generally, our beta findings may be related to proposals linking lower beta power to a change in the current cognitive set, where incoherent sentences might result in local sentence-level disruptions in syntactic and/or semantic processing, which in turn provide the language comprehension system with cues indicating that the current mode of processing needs to change. Finally, more attention appears to have been allocated to individual words in each sentence for coherent stories, as indicated by an enhanced visual N1 ERP component. We conclude that discourse-level semantic information is used during local sentence-level language comprehension, and has an effect on electrophysiological signatures of brain activity related to such processing.

Note

1. Ratings task results presented in Table 2 are those for the items included in the final stimulus set. The original rating task (see Hoffmann, 2011) included a larger set of stimuli and items not exhibiting the coherence breakdown at sentences 2–4, as indicated by the participant ratings, were excluded from the final set of stimuli used in our experiment.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is partly supported by a Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) VIDI grant to J.S. [grant number 864.14.011], and an IMPRS Ph.D. fellowship from the Max Planck Society to A.G.L.

ORCID

Ashley Glen Lewis  <http://orcid.org/0000-0003-4737-2525>
Jan-Mathijs Schoffelen  <http://orcid.org/0000-0003-0923-6610>

References

- Bastiaansen, M. C. M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research*, 159(6), 179–196. doi:10.1016/S0079-6123(06)59012-0
- Bastiaansen, M. C. M., & Hagoort, P. (2015). Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *Journal of Cognitive Neuroscience*, 27(11), 2095–2107. doi:10.1162/jocn_a_00829
- Bastiaansen, M. C. M., Magyari, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*, 22(7), 1333–1347. doi:10.1162/jocn.2009.21283
- Bastiaansen, M. C. M., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: Oscillatory neuronal dynamics. In S. Luck & E. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 31–49). New York, NY: Oxford University Press.
- Davidson, D. J., & Indefrey, P. (2007). An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Research*, 1158, 81–92. doi:10.1016/j.brainres.2007.04.082
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121. doi:10.1038/nn1504
- Engel, A. K., & Fries, P. (2010). Beta-band oscillations – signalling the status quo? *Current Opinion in Neurobiology*, 20(2), 156–165. doi:10.1016/j.conb.2010.02.015
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84. doi:10.1016/S1364-6613(00)01839-8
- Gray, C., König, P., Engel, A., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Letters to Nature*, 338, 334–337.
- Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 694–699. doi:10.1073/pnas.98.2.694
- Haarmann, H. I., Cameron, K. a., & Ruchkin, D. S. (2002). Neural synchronization mediates on-line sentence processing: EEG coherence evidence from filler-gap constructions. *Psychophysiology*, 39(6), 820–825.

- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9(9), 416–423. doi:10.1016/j.tics.2005.07.004
- Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, 4, 416. doi:10.3389/fpsyg.2013.00416
- Hagoort, P., & van Berkum, J. J. A. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1481), 801–811. doi:10.1098/rstb.2007.2089
- Hald, L. A., Bastiaansen, M. C. M., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, 96(1), 90–105. doi:10.1016/j.bandl.2005.06.007
- Hanslmayr, S., Staudigl, T., & Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: The information via desynchronization hypothesis. *Frontiers in Human Neuroscience*, 6(74), 1–12. doi:10.3389/fnhum.2012.00074
- Hipp, J. F., & Siegel, M. (2013). Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. *Frontiers in Human Neuroscience*, 7, 338. doi:10.3389/fnhum.2013.00338
- Hoffmann, C. (2011). *Do neuronal oscillations reflect semantic coherence in short discourses?* (Unpublished master's thesis). Nijmegen, Radboud University (Supervisors: Bastiaansen, M. C. M. & Schriefers, H.).
- Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 111(10), 1745–1758. doi:10.1016/S1388-2457(00)00386-2
- Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, 9(11), 512–518.
- Kielar, A., Meltzer, J. A., Moreno, S., Alain, C., & Bialystok, E. (2014). Oscillatory responses to semantic and syntactic violations. *Journal of Cognitive Neuroscience*, 26(12), 2840–2862. doi:10.1162/jocn_a_00670
- Kielar, A., Panamsky, L., Links, K. A., & Meltzer, J. A. (2014). Localization of electrophysiological responses to semantic and syntactic anomalies in language comprehension with MEG. *NeuroImage*, 105, 507–524. doi:10.1016/j.neuroimage.2014.11.016
- König, P., & Schillen, T. (1991). Stimulus-dependent assembly formation of oscillatory responses: I. Synchronization. *Neural Computation*, 3, 155–166. doi:10.1162/neco.1991.3.2.155
- Koornneef, A. W., & Sanders, T. J. M. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, 28(8), 1169–1206. doi:10.1080/01690965.2012.699076
- Kopell, N., Whittington, M. A., & Kramer, M. A. (2011). Neuronal assembly dynamics in the beta1 frequency range permits short-term memory. *Proceedings of the National Academy of Sciences of the United States of America*, 108(9), 3779–3784. doi:10.1073/pnas.1019676108
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647. doi:10.1146/annurev.psych.093008.131123
- Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163. doi:10.1038/307161a0
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2–3), 259–284. doi:10.1080/01638539809545028
- Lewis, A. G. (2012). *The role of oscillatory neuronal dynamics in discourse-level semantic comprehension* (Unpublished master's thesis). Nijmegen, Radboud University. (Supervisors: Bastiaansen, M. C. M. & Schriefers, H.).
- Lewis, A. G., Wang, L., & Bastiaansen, M. C. M. (2015). Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction? *Brain and Language*, 148, 51–63. doi:10.1016/j.bandl.2015.01.003
- Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, 8(5), 204–210. doi:10.1016/j.tics.2004.03.008
- Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences of the United States of America*, 94(20), 10979–10984. doi:10.1073/pnas.94.20.10979
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024
- Meyer, L., Obleser, J., & Friederici, A. D. (2013). Left parietal alpha enhancement during working memory-intensive sentence processing. *Cortex*, 49(3), 711–721. doi:10.1016/j.cortex.2012.03.006
- Meyer, L., Obleser, J., Kiebel, S. J., & Friederici, A. D. (2012). Spatiotemporal dynamics of argument retrieval and reordering: An fMRI and EEG study on sentence processing. *Frontiers in Psychology*, 3, 1–13. doi:10.3389/fpsyg.2012.00523
- Mitra, P. P., & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical Journal*, 76(2), 691–708. doi:10.1016/S0006-3495(99)77236-X
- Moreno, I., de Vega, M., León, I., Bastiaansen, M. C. M., Lewis, A. G., & Magyari, L. (2015). Brain dynamics in the comprehension of action-related language. A time-frequency analysis of mu rhythms. *NeuroImage*, 109, 50–62. doi:10.1016/j.neuroimage.2015.01.018
- Nieuwland, M. S., & van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7), 1098–1111. doi:10.1162/jocn.2006.18.7.1098
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 1–9. doi:10.1155/2011/156869
- Paczynski, M., & Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb-argument processing. *Language and Cognitive Processes*, 26(9), 1402–1456. doi:10.1080/01690965.2011.580143
- Peña, M., & Melloni, L. (2012). Brain oscillations during spoken sentence processing. *Journal of Cognitive Neuroscience*, 24(5), 1149–1164. doi:10.1162/jocn_a_00144
- Penolazzi, B., Angrilli, A., & Job, R. (2009). Gamma EEG activity induced by semantic violation during sentence reading. *Neuroscience Letters*, 465(1), 74–78. doi:10.1016/j.neulet.2009.08.065

- Pfurtscheller, G., & Lopes da Silva, F. (1999). *Event-related desynchronization. Handbook of electroencephalography and clinical neurophysiology, revised series (Vol. 6)*. Amsterdam: Elsevier B.V.
- Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, 25(6), 762–776. doi:10.1162/jocn_a_00337
- Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing. *Annual Review of Physiology*, 55, 349–374.
- St. George, M., Mannes, S., & Hoffnagel, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, 6(1), 70–83. doi:10.1162/jocn.1994.6.1.70
- Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3(4), 151–162. doi:10.1016/S1364-6613(99)01299-1
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *The Journal of Neuroscience*, 16(13), 4240–4249.
- Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., ... Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: Function, structure and behaviour in patients with left hemisphere damage. *Brain*, 134(2), 415–431. doi:10.1093/brain/awq369
- van Berkum, J. J. A. (2012). The electrophysiology of discourse and conversation. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 589–612). Cambridge: Cambridge University Press.
- van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, 1146(1), 158–171. doi:10.1016/j.brainres.2006.06.091
- van Berkum, J. J. A., Zwitserlood, P., Bastiaansen, M. C. M., Brown, C., & Hagoort, P. (2004). *So who's "he" anyway? Differential ERP and ERSP effects of referential success, ambiguity and failure during spoken language comprehension*. Paper presented at the Annual Meeting of the Cognitive Neuroscience Society, San Francisco.
- van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22(1), 67–82. doi:10.1162/jocn.2008.21170
- van Elk, M., van Schie, H. T., Zwaan, R. A., & Bekkering, H. (2010). The functional role of motor activation in language processing: Motor cortical oscillations support lexical-semantic retrieval. *NeuroImage*, 50(2), 665–677. doi:10.1016/j.neuroimage.2009.12.123
- Varela, F., Lachaux, J., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2, 229–239. doi:10.1038/35067550
- Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, 37(2), 190–203.
- Wang, L., Jensen, O., van den Brink, D., Weder, N., Schoffelen, J., Magyari, L., ... Bastiaansen, M. C. M. (2012). Beta oscillations relate to the N400 m during language comprehension. *Human Brain Mapping*, 33(12), 2898–2912. doi:10.1002/hbm.21410
- Wang, L., Zhu, Z., & Bastiaansen, M. (2012). Integration or predictability? A further specification of the functional role of gamma oscillations in language comprehension. *Frontiers in Psychology*, 3(187). doi:10.3389/fpsyg.2012.00187
- Weiss, S., & Mueller, H. M. (2003). The contribution of EEG coherence to the investigation of language. *Brain and Language*, 85(2), 325–343. doi:10.1016/S0093-934X(03)00067-1
- Weiss, S., & Mueller, H. M. (2012). "Too many betas do not spoil the broth": The role of beta brain oscillations in language processing. *Frontiers in Psychology*, 3(201). doi:10.3389/fpsyg.2012.00201
- Weiss, S., Mueller, H. M., Schack, B., King, J. W., Kutas, M., & Rappelsberger, P. (2005). Increased neuronal communication accompanying sentence comprehension. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 57(2), 129–141. doi:10.1016/j.ijpsycho.2005.03.013
- Womelsdorf, T., Fries, P., Mitra, P. P., & Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature*, 439(7077), 733–736. doi:10.1038/nature04258