An Electrophysiological Perspective on the Resolution of Anaphoric Dependencies: Evidence from Event-Related Potentials and Neural Oscillations

Cas Coopmans[1,2]
Supervisor: Mante Nieuwland[1,2]

[1]*Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands*
[2]*Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands*

In this study, we used electroencephalography to investigate the effects of givenness and discourse coherence of proper names on the electrophysiological correlates of anaphoric reference resolution. Participants read two-sentences mini-discourses in which repeated and new proper names were coherent or incoherent with the preceding discourse. Preregistered analyses revealed effects of givenness and coherence in both event-related potentials (ERPs) and neural oscillatory dynamics. In comparison with new names, repeated names elicited a reduced N400 and reduced Late Positive Component, and an increase in theta-band (4-7 Hz) synchronization. Discourse-coherent names elicited an increase in gamma-band (60-80 Hz) synchronization in comparison to discourse-incoherent ones, while no effects of discourse were observed in the ERP signal. We additionally observed an Nref ERP effect for new proper names that could not be properly linked to a discourse referent. We argue that the observed theta old/new effect reflects the reactivation of a representation held in working memory, and interpret the gamma-band effect in terms of successful semantic integration of coherent proper names. These results demonstrate that discourse-level anaphoric reference can be studied through neural oscillations and further establish the role of memory mechanisms in discourse-level language comprehension.

Keywords: Proper names, discourse comprehension, reference resolution, oscillations, ERPs, theta, gamma

---

Corresponding author: Cas Coopmans ; **E-mail**:cas.coopmans@mpi.nl

A cognitive perspective on anaphoric reference

Anaphoric dependency refers to the referential relationship between two linguistic representations that are separated from each other in time. This property of human language poses a real challenge to the comprehension system, because successful reference resolution requires multiple cognitive systems to work in concert. Consider the following discourse context: 'Lionel Messi and Cristiano Ronaldo are two of the world's greatest football players. Unsurprisingly, last year's Ballon d'Or was won by Ronaldo'. In order to understand that Ronaldo in the second sentence refers to the Portuguese Cristiano Ronaldo and not to the resigned Brazilian football player who happens to bear the same name, one has to establish a link between the anaphor 'Ronaldo' in the second sentence and its antecedent 'Cristiano Ronaldo' in the preceding discourse. In addition, in order to evaluate the correctness of this statement, the anaphor has to be integrated into the semantic representation of this sentence. In other words, in order to understand anaphoric dependencies, the language processor has to be able to distinguish between what is new and what has been referred to before, a distinction termed givenness, and to build a coherent and meaningful discourse representation on the basis of the meaning of the individual words. A recent proposal by Nieuwland and Martin (2017) links these two processes (i.e., memory retrieval and semantic integration), assumed to be core components underlying anaphor resolution, to increased oscillatory synchronization in two ranges of the gamma band. Nieuwland and Martin argue that anaphor comprehension draws on an interaction between the brain's recognition memory network and the language system. This specific proposal forms part of a bigger enterprise that aims to answer the question whether the neural mechanisms classically thought to be involved in memory functioning might also underlie online language processing (Covington & Duff, 2016; Duff & Brown-Schmidt, 2012; Piai et al., 2016).

In an attempt to test Nieuwland and Martin's proposal, we performed an electroencephalography (EEG) study to further investigate the effects of givenness and coherence on neural oscillations. In order to study the effects of givenness, we utilized a key attribute of proper names, which is that they can both establish reference by introducing a new referent into the discourse and maintain reference by referring to a given discourse referent. These proper names were embedded in two-sentence mini-discourses, in which we additionally manipulated the coherence of the proper names with respect to the preceding discourse. By looking at how these effects play out in electrophysiological brain recordings, we aim to contribute to the emerging view on the role of memory processes in language comprehension.

An electrophysiological perspective on anaphoric reference

One prominent memory-based approach to linguistic dependency resolution is known as the cue-based retrieval framework (Martin, 2016; McElree, 2000; McElree, Foraker & Dyer, 2003). The cue-based retrieval framework argues that the memory system subserving sentence comprehension has a content-addressable architecture, in which access and retrieval of content-addressable memory representations is cue-induced and direct, without a search through irrelevant representations. Retrieval success is dependent on the match between the cue that triggered retrieval and the target representation (McElree, 2006). Specific to anaphor comprehension, this equates to identifying the referent of an anaphor by evaluating the overlap between the representation of the anaphor that triggered retrieval and the representation of the antecedent held in working memory.

Recent studies have demonstrated the potential of the cue-based retrieval framework to account for the electrophysiological correlates of anaphoric reference by using event-related potentials (ERPs) to study cue-based retrieval during online language comprehension (e.g., Martin, Nieuwland & Carreiras, 2012, 2014). ERPs are voltage fluctuations in the electroencephalogram that are evoked by an event, such as an external stimulus. Comparing the average ERP signal between conditions reveals ERP effects, which, by virtue of their multidimensional nature (i.e., polarity, amplitude, time course, scalp topography), can provide both quantitative and qualitative information about the cognitive event at hand (Van Berkum, 2004). The downside of ERPs is that, due to averaging, any event-related response that is nonstationary (i.e., has a jitter in phase and/or latency) will be canceled (Tallon-Baudry & Bertrand, 2000). ERPs therefore only capture neural activity that is strictly time- and phase-locked to the stimulus event.

Non-phase-locked neural activity, known as event-related neural oscillations, are event-induced modulations of ongoing rhythmic patterns of neural activity. Because oscillations are always present, the phase of the oscillatory signal at the onset of the event is variable, meaning that event-induced changes in the oscillatory EEG activity are not strictly phase-locked (Bastiaansen, Mazaheri & Jensen, 2012). This non-phase-locked activity can be visualized by a time-frequency approach that computes the power of the activity at each individual frequency prior to averaging. As power (amplitude squared) is a positive measure, averaging the time-frequency signal over multiple trials leaves in both phase-locked and non-phase-locked activity. The benefit of studying both ERPs and neural oscillations is that they, because they reflect dissociable aspects of the same neuronal activity, might provide a complementary view on the neural activity related to cognitive processes (Davidson & Indefrey, 2007; Kielar, Panamsky, Links & Melzer, 2015). In the following sections, we will review the electrophysiological literature on memory reactivation and semantic integration during language comprehension.

Event-related potentials and anaphoric reference

Givenness

Behavioral studies have established that word repetition leads to repetition priming: repeated words are easier to process than new words, as demonstrated via lexical decision tasks (e.g., Scarborough, Cortese & Scarborough, 1977). In EEG studies of recognition memory, this repetition priming effect is known as the ERP old/new effect, which often takes the form of a reduction of the amplitude of the N400 component and a subsequent Late Positive Component (LPC) for repeated compared to new words (Rugg, 1985, 1990; Van Strien, Hagenbeek, Stam, Rombouts & Barkhof, 2005; Rugg & Curran, 2007). The N400 component is a negative deflection peaking between 300-500 ms after the onset of each content word and is largest over centro-parietal electrodes (the difference between two components constitutes the N400 effect; Kutas & Hillyard, 1980; Kutas & Federmeier, 2011), while the LPC has a parietal distribution, begins approximately around 500 ms after word onset and can last until up to 1000 ms (for a review, see Van Petten & Luka, 2012). In sentence and discourse context, where repetition is a natural means to establish co-reference, repeated words also elicit a reduced N400 component (Van Petten, Kutas, Kluender, Mitchiner & McIsaac, 1991; Streb, Henninghause & Rösler, 2004; Camblin, Ledoux, Boudewyn, Gordon & Swaab, 2007a; Ledoux, Gordon, Camblin & Swaab, 2007; Almor, Nair, Boiteau & Vendemia, 2017), but new words elicit a larger LPC than repeated words (Burkhardt, 2006, 2007; Kaan, Dallas & Barkley, 2007; Schumacher, 2009; Wang & Schumacher, 2013). The attenuated N400 in response to a repeated word (i.e., anaphor) seems to reflect the ease with which its referent (i.e., antecedent) is identified within the preceding discourse, while the LPC for new words indexes augmentation of the current discourse model with a new entity (Burkhardt, 2006; Schumacher, 2009; Schumacher & Hung, 2012; Wang & Schumacher, 2013). Evidence for this two-process interpretation was provided by Burkhardt (2006), who demonstrated that inferentially-bridged noun phrases, such as 'the conductor' in (1), first pattern with repeated words (2) in terms of a reduced N400 and then with new words (3) in terms of an enhanced LPC.

1) Tobias visited a *concert* in Berlin. He said that **the conductor** was very impressive.

2) Tobias visited a *conductor* in Berlin. He said that **the conductor** was very impressive.

3) Tobias talked to *Nina*. He said that **the conductor** was very impressive.

Although 'the conductor' in (1) might rather effortlessly be anchored to the discourse based on inferential knowledge (i.e., *concert* implies *conductor*), explaining the reduced N400, it will still have to be introduced into the discourse model as an independent representation, explaining the enhanced LPC (Burkhardt, 2006).

Discourse coherence

In line with these findings, the amplitude of the N400 has been argued to reflect the ease or difficulty with which conceptual knowledge associated with words or names is retrieved (Kutas & Federmeier, 2011; Van Berkum, 2012). This process can be modulated by context, but only to the comprehender's benefit. That is, while a coherent context can reduce the N400 amplitude by facilitating ease of retrieval, no additional retrieval cost (i.e., enhanced N400) seems to be incurred from an incoherent context (Hagoort & Van Berkum, 2007; Kutas & Federmeier, 2011; Van Petten & Luka, 2012). The effects of discourse on semantic retrieval have been addressed by numerous studies, which all showed that the N400 is reduced for words that are coherent with respect to the preceding discourse compared to discourse-incoherent words (Camblin, Gordon & Swaab, 2007b; Filik & Leutholdt, 2008; Nieuwland & Van

Berkum, 2006a; Salmon & Pratt, 2002; St. George, Mannes & Hoffman, 1994; Van Berkum, Hagoort & Brown, 1999a; Van Berkum, Zwitserlood, Hagoort & Brown, 2003). An important finding of these studies was that the discourse-level and sentence-level N400 were identical in terms of time course, morphology and scalp distribution (Van Berkum et al., 1999a, 2003; Salmon & Pratt, 2002; Nieuwland & Van Berkum, 2006a), indicating that the processes responsible for the N400 do not seem to be sensitive to where the semantic constraints come from (Van Berkum, 2004, 2012).

With particular relevance to the current study are the findings from a study by Wang and Yang (2013), who demonstrated the beneficial effects of context on the retrieval of proper names from working memory. They set up a two-sentence discourse context in which two proper names were introduced and described with contrastive characteristics (e.g., 'John is a *singer*, Peter is an *actor*'). In the third sentence, the interpretation of the critical proper name was either coherent or incoherent with respect to the preceding discourse (e.g., 'a film producer/music producer came to *Peter*'). Compared to discourse-coherent names, incoherent proper names elicited a widely distributed N400 effect and an additional P600 effect, showing that the learned meaning of previously unknown proper names can easily and rapidly be reactivated from working memory and integrated into the context.

Givenness in discourse

Although repetition in language context seems to show a similar N400 profile as repetition in word lists, the discourse context has been shown to modulate this effect. That is, when repeated proper names refer to antecedents that are in discourse focus, as in (4), they elicit a larger N400 than repeated proper names that refer to an antecedent that is not prominent in the discourse model, as in (5) (Swaab, Camblin & Gordon, 2004).

4) At the office *Daniel* moved the cabinet because **Daniel** needed room for a desk.

5) At the office *Daniel and Amanda* moved the cabinet because **Daniel** needed room for a desk.

This effect is an electrophysiological manifestation of the so-called repeated name penalty (Gordon, Hendrick, Ledoux & Yang, 1999), and has been observed in both reading (Swaab et al., 2004; Ledoux et al., 2007) and listening to fully connected natural speech (Camblin et al, 2007a). Notably, the N400 elicited by repeated name anaphors that co-refer with a discourse-prominent antecedent is comparable to the N400 elicited by mentioning a new name, suggesting that the repeated-name penalty reflects the effect of discourse prominence overriding otherwise facilitatory effects of repetition (Camblin et al., 2007a).

A referentially induced negativity

The ERP effect most prominently associated with referential processing is a sustained negativity with an anterior distribution (for a review, see Van Berkum, Koornneef, Otten & Nieuwland, 2007). Van Berkum, Brown and Hagoort (1999b) observed that referentially ambiguous noun phrases, such as 'the girl' in a discourse that involved two girls, elicited a rapidly (~300 ms after noun onset) emerging and relatively long-lasting negativity with a predominantly frontal distribution. This sustained frontal negativity in response to written referential ambiguity was subsequently replicated in the auditory modality, using the same mini-discourses presented as naturally produced connected speech (Van Berkum, Brown, Hagoort & Zwitserlood, 2003). Later studies by Van Berkum, Nieuwland and colleagues showed that this referentially induced frontal negativity, termed Nref, is reliably elicited in the comparison between unambiguous and ambiguous anaphors, such as 'he' in 'David told Peter that he …' where both 'David' and 'Peter' are equally plausible antecedents (Nieuwland, 2014; Nieuwland, Otten & Van Berkum, 2007a; Nieuwland & Van Berkum, 2006b; Van Berkum, Zwitserlood, Bastiaansen, Brown & Hagoort, 2004). Note that the results of these studies do not imply that "the processes directly responsible for the negativity here must be uniquely tied to resolving referential ambiguity" (Van Berkum, 2009, p. 301), but merely show that "the processing consequences of referential ambiguity quite consistently show up as sustained frontal negativities" (Van Berkum, 2012, p. 600).

In line with these latter statements, a recent ERP study found an Nref-like effect in response to the resolution of anaphors that were referentially unambiguous (Barkley, Kluender & Kutas, 2015). Barkley and colleagues presented participants pronouns and proper names that either did (6) or did not (7) have an antecedent earlier in the sentence.

6) After a covert mission that deployed *Will*$_i$/*him*$_j$ for nine terrible months, **he**$_i$**/Will**$_j$ longed for home.

7) After a covert mission that required deployment for nine terrible months, **he/Will** longed for home.

In comparison with pronouns without antecedents, pronouns with antecedents elicited a large negativity with an anterior distribution. No difference was observed between proper names with and without antecedents. These findings were interpreted in terms of the cue-based retrieval properties of certain anaphoric forms (McElree et al., 2003). According to the authors, pronouns contain a [+ antecedent] feature that cues a process they call ‘back association’, which involves the reactivation of an already encoded antecedent in order to allow the establishment of a referential dependency. Proper names, in contrast, are known to be primarily used to introduce entities into the discourse rather than to maintain reference (Gordon & Hendrick, 1998; Gordon et al., 1999). Accordingly, proper names were argued not have the [+ antecedent] feature and therefore do not trigger back association (Barkley et al., 2015).

The fact that several ERP effects are seen in studies aimed at identifying the electrophysiological correlates of anaphoric reference indicates that it does not rely on a single process. Rather, it relies on the cooperation of multiple cognitive operations, among others retrieval from working memory and semantic integration. ERPs are inherently limited in what they can tell us about the interaction between these operations. The study of neural oscillations, however, might allow us to gain more insight into not only which cognitive components are involved in anaphor comprehension (e.g., via investigation of local oscillatory power) but also how these components work together in order to establish a coherent discourse representation (e.g., via studying global phase coherence).

Neural oscillations and anaphoric reference

Different aspects of language comprehension have been related to different oscillatory frequency bands. Memory reactivation and semantic integration have most prominently been associated with oscillations in the theta and gamma frequency bands (for a recent review, see Meyer, 2017).

Theta oscillations

With particular relevance for the current study are two lines of research that have observed a relationship between retrieval from both long-term memory and working memory and activity in the theta band (4-7 Hz). First, theta-band oscillations have been related to the ease with which lexico-semantic information is retrieved from long-term memory, as evidenced by larger theta power for semantically rich compared to semantically lean words (Bastiaansen, Van der Linden, Ter Keurs, Dijkstra & Hagoort, 2005), larger theta power for semantically anomalous compared to semantically coherent words (Davidson & Indefrey, 2007; Hald, Bastiaansen & Hagoort, 2006; Wang, Zhu & Bastiaansen, 2012) and differences in the scalp distribution of theta power as a function of the modality-specific properties of words (Bastiaansen, Oostenveld, Jensen & Hagoort, 2008). Second, in the recognition memory literature, theta is related to the recognition and retrieval of successfully remembered probes (see Nyhuis & Curran, 2010 for a review). These studies either employ a study-test or a continuous recognition paradigm. In the former, participants are asked to study a list of items, often words, and after some time (sometimes on a different day) are given another list of items and are asked to indicate whether these items had been studied before (‘old’) or not (‘new’). In continuous recognition paradigms, participants walk through a list of items and have to judge continuously for each individual item whether it has been presented in the list before or not. Correctly remembered old items reliably elicited larger theta-band synchronization than correctly recognized new items, both in the study-test paradigm (Chen & Caplan, 2016; Jacobs, Hwang, Curran & Kahana, 2006; Klimesch, Doppelmayr, Schimke & Ripper, 1997; Klimesch, Doppelmayr, Schwaiger, Winkler & Gruber, 2000; Osipova et al., 2006) and in the continuous recognition task (Burgess & Gruzelier, 1997, 2000; Van Strien et al., 2005; Van Strien, Verkoeijen, Van der Meer & Franken, 2007). These theta oscillations have therefore been linked to the process of matching the probe to representations in working memory, where old and new items are differentiated (Jacobs et al., 2006; Chen & Caplan, 2016).

One could argue that, despite potential differences in task-induced strategies, the retrieval demands of continuous recognition and anaphor comprehension are conceptually similar. In both cases, an item (probe, anaphor) triggers the retrieval of a recently encoded item (target, antecedent) that is held in working memory. In addition, in both cases the items are separated in time and intervened with other items that might interfere with retrieval success. The core computational operations involved in recognition memory and anaphor comprehension (i.e., distinguishing old from new information; matching input to a representation in working memory) are thus conceptually related and might even be shared (Covington & Duff, 2016; Martin, 2016; Nieuwland & Martin, 2017).

Gamma oscillations

Oscillations in the gamma-band (>30 Hz), as found in language experiments, have been related to semantic integration processes (called semantic unification; Bastiaansen & Hagoort, 2006; Hagoort, Willems & Baggio, 2009). These experiments show that whenever semantic unification is successful, gamma-band power is increased. For instance, gamma-band synchronization is increased for referentially coherent pronouns relative to pronouns that are referentially ambiguous or do not have a proper referent (Van Berkum, Zwitserlood, Bastiaansen, Brown & Hagoort, 2004), for semantically coherent sentences compared to syntactic prose (i.e., syntactically correct but semantically anomalous sentences; Bastiaansen & Hagoort, 2015), and for semantically coherent compared to anomalous words (Hald et al., 2006; Penolazzi, Angrili & Job, 2009), the latter effect being modulated by the semantic relatedness between the anomalous word and the sentence context (Rommers, Bastiaansen & Dijkstra, 2013).

In a recent paper, Nieuwland and Martin (2017) reported time-frequency analyses of four previous ERP studies on referential processing. All four studies compared referentially coherent anaphors to referentially incoherent ones (ambiguous or problematic; e.g., 'Peter thought that *he/she* would win the race'). Two bursts of gamma-band power were observed in response to referentially coherent anaphors: an increase in 35-45 Hz ('low') gamma-band power between 400 and 600 ms and another gamma-band increase of 60-85 Hz ('high gamma') between 500 and 1000 ms. Source localization revealed a strong source for the low gamma effect in the left posterior parietal cortex (LPPC), while the high gamma effect was localized to the left inferior frontal-temporal cortex, including the left inferior frontal gyrus (LIFG). Largely because the LPPC has been shown to differentiate old from new information (Gonzalez et al., 2015) and has been related to successful retrieval from episodic memory (for a review, see Wagner, Shannon, Kahn & Buckner, 2005), the low gamma effect was taken to reflect the reactivation of the antecedent by the brain's recognition memory network. As the LIFG is known to be involved in sentence-level semantic unification (Hagoort, 2005; Hagoort & Indefrey, 2014), the high gamma effect was argued to reflect the integration of the reactivated antecedent into the semantic representation of the sentence.

Notably, Nieuwland and Martin (2017) acknowledge that these interpretations face several challenges. First of all, although the two gamma effects were assigned different functional interpretations, it is questionable whether they are truly distinct. Secondly, the low gamma effect was not found in all of their experiments, and it shows up in a frequency range that is not commonly associated with successful language comprehension, nor with successful memory retrieval. It is therefore not yet clear whether these gamma effects reflect the workings of a recognition memory network or whether they are more specifically related to processes involved in resolving referential ambiguity. Third, memory processes have been associated most prominently with the oscillations in the theta band, but it is unclear how these relate to the memory mechanisms that are relevant for anaphor comprehension. We aim to address these questions in a dedicated EEG experiment that focuses on the effects of givenness and coherence of proper names in discourse.

Present study

The present EEG study uses ERPs and neural oscillations to investigate the involvement of memory retrieval and semantic integration in anaphor comprehension. We use two-sentence mini-discourses in which the variables givenness (old/anaphor vs. new/non-anaphor) and coherence (discourse-coherent vs. discourse-incoherent) are orthogonally manipulated in a two-by-two design. An example of a stimulus item is given in Table 1. The first sentence of each mini-discourse introduces two entities by name (e.g., John and Peter). In the second sentence, a critical proper name was used either anaphorically, referring back to one of the two previously mentioned entities ('old'; e.g., John when John and Peter are introduced), or non-anaphorically, introducing a new entity into the discourse model ('new'; e.g., John when David and Peter are introduced). In addition, the interpretation of the critical proper name in the second sentence is either coherent or incoherent with respect to the preceding discourse. In order to examine the brain's response to a new, non-anaphoric proper name when there are no specific referents in the preceding discourse, we added a neutral condition in which a new proper name could be anaphorically linked to a non-specific, generic antecedent in the preceding

Table 1.
Example of one stimulus item, containing all five conditions of an original Dutch two-sentence mini-discourse. Approximate English translations of each sentence are provided below. The critical proper names (CW) are in bold. Characteristic information that was manipulated is underlined.

	Sentence 1	**Sentence 2**
Old-coherent	**Jan** en Peter zijn de beste spelers uit het voetbalelftal. *(**John** and Peter are the best players in the football team.)*	De topscorer van het team was **Jan** met dertig doelpunten in totaal. *(The top scorer of the team was **John** with thirty goals in total.)*
Old-incoherent	**Jan** en Peter zijn de slechtste spelers uit het voetbalelftal. *(**John** and Peter are the worst players in the football team.)*	
New-coherent	David en Peter zijn de slechtste spelers uit het voetbalelftal. *(David and Peter are the worst players in the football team.)*	
New-incoherent	David en Peter zijn de beste spelers uit het voetbalelftal. *(David and Peter are the best players in the football team.)*	
New-neutral	De spelers in het voetbalelftal zijn erg goed. *(The players in the football team are very good.)*	

sentence (e.g., the players in the football team), and is neutral with respect to the coherence manipulation. This additionally allowed us to investigate Barkley et al.'s (2015) claim that proper names do not trigger back association. To be specific, if proper names do not trigger back association, no differences should be observed between new names for which a (generic) discourse referent is available (new-neutral) and new names that do not have an antecedent and therefore must introduce an entity into the discourse (new-coherent).

Hypotheses

For the ERP analysis, we expect that name repetition would elicit a biphasic ERP pattern: compared to repeated names, new names are expected to elicit a larger N400 and a subsequent LPC (e.g., Burkhardt, 2006). Similar to Wang and Yang (2013), we expect to observe a smaller N400 for discourse-coherent than discourse-incoherent proper names. A possible interaction effect might reveal itself in two ways. First, an incoherent discourse context could override the facilitatory effects of repetition (e.g., Camblin et al., 2007a; Ledoux et al., 2007). As a result, there will be no attenuation of the N400 for old-incoherent proper names. Alternatively, under the assumption that a possible N400 in our experiment reflects reactivation of an item in working memory, we expect the N400 in response to new names to be insensitive to discourse coherence. That is, a coherent discourse cannot not facilitate the reactivation of a new name, simply because a new name does not yet have a representation in working memory. Concerning the new-neutral condition, if proper names trigger back association, we expect an Nref effect for new-neutral compared to old-coherent proper names. No difference is expected between new-neutral and new-coherent, as both conditions contain a reference group to which the new proper name can be linked. However, if proper names do not trigger back association, as assumed by Barkley et al. (2015), no Nref effect is expected between new-neutral proper names and the coherent conditions.

We investigated oscillatory dynamics in three frequency ranges: theta (4-7 Hz), low gamma (34-45 Hz) and high gamma (60-80 Hz). Based on the observations of an increase in theta-band synchronization for old compared to new words (Burgess & Gruzelier, 1997, 2000; Klimesch et al., 1997, 2000), we tentatively hypothesize that increased theta-band synchronization reflects retrieval from working memory. These reactivation processes might also be linked to the gamma frequency band, as suggested by Nieuwland and Martin (2017). We therefore predict an increase in both theta- and low gamma-band power in response to old compared to new proper names. In line with the semantic unification literature (Bastiaansen & Hagoort, 2006, 2015), we expect to see an increase in high gamma-band power for discourse-coherent compared to discourse-incoherent proper names.

Methods

Preregistration

The design and settings of our analysis procedures (i.e., preprocessing, time-frequency analysis, and statistical analysis) have all been preregistered at Open Science Framework[1]. All non-preregistered analyses are exploratory and will explicitly be referred to as such.

Participants

A total sample size of 40 participants, after exclusions, was preregistered. In total, 45 native speakers of Dutch participated in the experiment. They were paid 18 euros for participation. All of them had normal or corrected-to-normal vision and none of them reported being dyslexic. Three participants had to be excluded after we found out that they were left-handed. Two additional participants had to be excluded from ERP analysis because too many trials had been rejected during ERP preprocessing. Similarly, after preprocessing the oscillatory data one participant ended up with too few trials and had to be excluded from the time-frequency analysis. Therefore, ERP analysis was done on 40 right-handed speakers of Dutch (30 females, average age: 23 years, age range: 19-33 years), while 41 right-handed speakers of Dutch (29 females, average age: 23 years, age range: 19-33 years) were included in the time-frequency analysis.

Materials

Experimental items

A total of 225 two-sentence mini-discourses were created. The second sentence of a given mini-discourse was identical for all conditions, while the first sentence varied between the conditions. The first sentence introduced two people by proper names within a conjoined noun phrase (e.g., John and Peter). This type of embedding has been shown to reduce the prominence of both referents, making subsequent use of a repeated name felicitous. The second half of the first sentence added characteristic information about these two people. This information could regard a social or personality characteristic (e.g., being friendly/unfriendly), a physical characteristic (e.g., being strong/weak) or a behavioral characteristic (e.g., getting good/bad grades). In addition, the first sentence always contained ‘a reference group’ to which both players belong (e.g., the football team). This was added to make sure the second sentence is not interpreted as referring to a situation in which only two entities are present. In the second sentence, a proper name (the critical word; CW) was used either anaphorically or non-anaphorically. Anaphoric use of the critical proper name represented the ‘old’ condition, because the exact same name had been introduced in the first sentence, whereas non-anaphoric use of the critical proper name represented the ‘new’ condition. The factor coherence (‘coherent’ vs. ‘incoherent’) was manipulated by using antonyms to denote the characteristic information across conditions within each item (e.g., friendly-unfriendly, getting good grades-getting bad grades). This made the interpretation of the critical proper name in the second sentence either coherent or incoherent with the characteristic information that was assigned to this proper name in the first sentence. In the fifth, new-neutral condition, the first sentence was kept semantically similar to the other conditions, with the exception that the proper names were removed and the reference group (e.g., the players in the football team) had become the subject of the sentence. This fifth condition was called new-neutral, because the proper name in the second sentence introduces a new entity (albeit interpretable as belonging to the reference group) and its interpretation in the second sentence is neutral with respect to the characteristic information provided in the first sentence. To avoid

---

[1] The project is called ‘Proper names in discourse’ and can be reached via https://osf.io/nbjfm/

sentence-final wrap-up effects contaminating the brain's response in the time window of interest, the CW was always followed by five words that ended the sentence without (anaphorically) referring to any of the entities.

The factors givenness (old, new) and coherence (coherent, incoherent) were orthogonally manipulated in a within-subjects two-by-two design, rendering the critical proper name in the second sentence old and coherent, new and coherent, new and incoherent or old and incoherent.

Approximately half (113) of the items only contained male names, the other half (112) contained only female names. All names were unambiguously either a male or female name. Each name was used only once in the entire experiment, meaning that we used 675 proper names (225 items x 3 names) in total.

To control for potential effects of order of mention, half of the anaphoric proper names in the old conditions referred to the first-mentioned name in the first sentence (John in the examples in Table 1), while the rest of the time it referred to the second-mentioned name in the first sentence (Peter in the examples in Table 1).

Filler items

We added 25 filler items to prevent participants from becoming too familiar with incoherent items. The first sentence of each filler item was very similar to the first sentence in the new-neutral condition. It contained a reference group (e.g., the candidates in the elections) with characteristic information (e.g., being popular). The second sentence was very similar to the second sentence in all experimental items, except for the fact that the proper name was replaced by a definite noun phrase referring to a specific person that belongs to the reference group (e.g., the politician). A translated example of a filler item: *The candidates in the elections are very popular. The majority of the people voted for the politician with the extraordinarily creative ideas.*

Comprehension questions

Eighty comprehension questions (50% with correct answer 'yes', 50% with correct answer 'no') were included to ensure that participants paid attention. These had to be answered by means of a button press. Questions were either about the general gist of the discourse story (50/80) or about specific entities in the stories (30/80). The average percentage of correctly answered questions was 92,4%. None of the participants scored below the preregistered cutoff of 70%.

Experimental lists

Five lists were made, each containing one condition of an experimental item. This was done to ensure that participants only saw one condition of each item. All lists had the same number of items of each condition. The filler items were the same for all lists. For each of the five lists, two versions were created by pseudorandomizing the order of the trials, with the only restriction that the same condition was never presented more than three times in a row. The participants were equally divided over all ten lists.

Procedure

Participants were comfortably seated in front of a computer screen in a fully shielded soundproof booth. After they had been informed about the procedure of the experiment, they were instructed to attentively and silently read sentences for comprehension and answer the comprehension questions. The sentences were presented in black letters (font Times New Roman, size 34) at the center of the screen, which had a light grey background.

Each trial started with a fixation cross (+) presented at the center of the screen. When participants pressed a button, the first sentence of each mini-discourse would be presented as a whole. After participants had carefully read the sentence, they pressed a button to start the presentation of the second sentence. This second sentence was presented word-by-word, which allowed us to control the time point at which the critical word was presented. Each word was presented for a duration of 400 ms, with the exception of the sentence-final word, which had a duration of 800 ms. The inter-word-interval was 200 ms in length. The sentence-final word was either followed by a fixation cross, indicating the start of the next trial, or a comprehension question. Participants were asked to minimize eye blinks and body movements during the word-by-word presentation of the second sentence.

The 250 (225 experimental + 25 filler) items were presented in five blocks of 50 items. Each block contained nine items of each condition and five filler items. Sixteen items per block were followed by a comprehension question. Participants were allowed to take short breaks between blocks. In total, the experiment lasted approximately 70 minutes (excluding EEG set-up time).

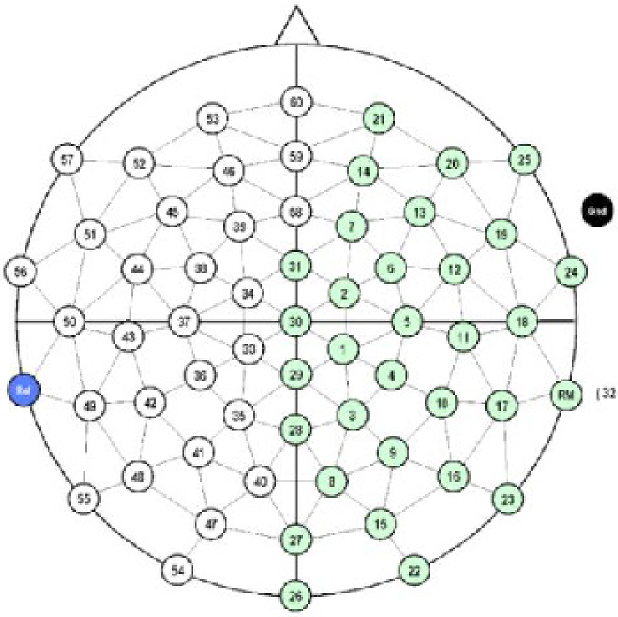

Fig. 1. Schematic representation of the 59-electrode array layout.

EEG recording and preprocessing

The electroencephalogram (EEG) was recorded using an MPI custom actiCAP 64-electrode montage (Brain Products, Munich, Germany), of which 59 electrodes were mounted in the electrode cap (see Figure 1). Horizontal eye movements (horizontal EOG) were recorded by one electrode placed on the outer canthus of the right eye, and eye blinks (vertical EOG) were recorded by two electrodes placed below both eyes. One electrode was placed on the right mastoid, the reference electrode was placed on the left mastoid and the ground was placed on the forehead. The EEG signal was amplified through BrainAmp DC amplifiers, referenced online to the left mastoid, sampled at 500 Hz and filtered with a passband of 0.016-249 Hz.

For ERPs, the data was band-pass filtered at 0.03-40 Hz (24 db/oct). Then, the data was re-referenced to the average of the left and right mastoid. Epochs were created ranging from -500 to 1500 ms relative to CW onset, and these were normalized to a 250 ms pre-CW baseline. Bad trials, containing low-frequency drifts, spikes or line noise were rejected through visual inspection. Independent Component Analysis (using ICA weights from a 1 Hz high-pass filtered version of the data) was used to filter artefacts resulting from eye movements and steady muscle activity. Epochs containing voltage values exceeding ± 90 µV were automatically rejected. On average, 13.5 ERP segments (average per condition ranged from 1.9 to 2.3) per subject were rejected. After preprocessing, two participants ended with less than 160 trials and were replaced.

For oscillations, the data was band-pass filtered at 0.1-100 Hz (24 db/oct) and re-referenced to the average of the left and right mastoid. Then, the data was segmented into epochs ranging from -1000 to 2500 ms relative to CW onset. Bad trials were again rejected through visual inspection. Independent Component Analysis (using ICA weights from a 1 Hz high-pass filtered version of the data) was used to filter artifacts resulting from eye movements and steady muscle activity. Because we did not apply baseline correction on the oscillatory data, using the preregistered automatic artifact rejection procedure based on an amplitude criterion (of ± 100 µV) would have excluded too many (good) trials. Therefore, we used a difference criterion that excluded segments in which the difference between the maximum and minimum voltage exceeded 200 µV. On average 12.1 segments (average per condition ranged from 2.1 to 2.5) per subject were rejected. One participant ended with less than 160 trials and was replaced.

Time-frequency (TF) analysis of oscillatory power was performed using the Fieldtrip toolbox (Oostenveld, Fries, Maris & Schoffelen, 2011). In order to find a right balance between time and frequency resolution, we performed time-frequency analysis in two different, but partially overlapping frequency ranges. For the low (2-30 Hz) frequency range, power was extracted from each individual frequency by moving a 400-ms Hanning window with ± 5 Hz spectral smoothing along the time axis in time steps of 10 ms. For the high (30-90 Hz) frequency range, we computed power changes with a multitaper approach (Mitra & Pesaran, 1999) based on discrete prolate spheroidal (Slepian) sequences as tapers, with a 400 ms time-smoothing and a ± 5-Hz spectral-smoothing window, in frequency steps of 2.5 Hz and time steps of 10 ms. On each individual trial, power changes in the post-CW interval were computed as a relative change from a baseline period ranging from -500 to -250 ms relative to CW onset. Per subject, we computed average power changes for each condition separately.

Statistical analysis

Statistical analysis of ERPs

We performed a linear mixed effects analysis (Baayen, Davidson & Bates, 2008) in R (R Core Team, 2018), using the lme4-package (Bates, Maechler & Bolker, 2012). The analyses were done separately for the N400, LPC and Nref regions-of-interest (ROIs).

At the N400 ROI, the dependent variable was, the average voltage across the centroparietal electrodes 35, 28, 3, 41, 40, 8, 9, 47, 27, 15 in a 300-500 ms window after CW onset (based on the spatial and temporal characteristics of the N400; e.g., Kutas & Federmeier, 2011). At the LPC ROI, the dependent variable was average across the same centroparietal electrodes 35, 28, 3, 41, 40, 8, 9, 47, 27, 15 in a 500-1000 ms window after CW onset (e.g., Van Petten & Luka, 2012). Dependent variables of the N400- and LPC ROIs were computed separately for each trial and each participant. The predictors coherence and givenness were deviation coded. We started with a full model that included the main effects of givenness (new, old) and coherence (coherence, incoherent), as well as their two-way interaction as fixed effects terms. Subject and item were included as random effects. Following Barr, Levy, Scheepers and Tily (2013), we attempted to use a maximal random effects structure by including the interaction between givenness and coherence as by-subject and by-item random slope. Because these models did not converge, we tested the same models without random correlations. As this still led to non-convergence, we removed the interaction term from the random slope and tested models with the same predictors and a by-participant and by-item random intercept only. Models are specified in the footnote[2]. In order to locate the model with the best fit, we started from a full model that included the interaction term and both main effects, and reduced its complexity stepwise, by first removing the interaction and then the main effects. Models were compared using R's `anova()` function, and *p*-values below α = .05 were treated as significant.

At the Nref ROI, the dependent variable was the average voltage across the frontal electrodes 53, 60, 21, 46, 59, 14, 39, 58, 7 in a 300-1500 ms window after CW onset (e.g., Van Berkum et al., 2007). In two separate analyses, we tested the effect of condition, where condition either had the levels 'new-neutral' and 'new-coherent' or 'new-neutral' and 'old-coherent'[3]. Subject and item were entered as random effects, and we included condition as by-subject and by-item random slope. We compared the models with and without condition using R's `anova()` function. *P*-values below α = .05 were treated as significant.

Statistical analysis of oscillatory power

We used cluster-based random permutation tests (Maris & Oostenveld, 2007) to compare differences in oscillatory power across conditions. This non-parametric statistical test deals with the multiple comparisons problem by statistically evaluating cluster-level activity rather than activity at individual data points, and is based on the fact that effects in electrophysiological data tends to be clustered in time, space and frequency (Maris, 2012). By evaluating the test statistic of the multidimensional cluster it retains statistical sensitivity while controlling the false alarm rate.

In brief, the cluster-based random permutation test works as follows: first, by means of a two-sided dependent samples t-test we performed the comparisons described below, yielding uncorrected p-values. Neighboring data triplets of electrode, time and frequency-band that exceeded a critical α-level of .05 were clustered. Clusters of activity were evaluated by comparing their test cluster-level statistic (sum of individual *t*-values) to a Monte-Carlo permutation distribution that was created by computing the largest cluster-level *t*-value on 1000 permutations of the same dataset. Clusters falling in the highest or lowest 2.5th percentile were considered significant. We used the correct-tail option that corrects *p*-values for doing a two-sided test, which allowed us to evaluate *p*-values at α = .05.

The following comparisons had been preregistered: a contrast between old (average of old-coherent and old-incoherent) and new (average of new-coherent and new-incoherent) in the 4-7 Hz

---

[2] N400/LPC:
Model1: N400/LPC ~ givenness*coherence + (1 | subject) + (1 | item)
Model2: N400/LPC ~ givenness + coherence + (1 | subject) + (1 | item)
Model3: N400/LPC ~ givenness + (1 | subject) + (1 | item)
Model4: N400/LPC ~ coherence + (1 | subject) + (1 | item)
Model5: N400/LPC ~ (1 | subject) + (1 | item)

[3] Nref:
Model1: Nref ~ condition + (condition|subject) + (condition|item)
Model2: Nref ~ (condition|subject) + (condition|item)

(theta) frequency range in a 0-1000 ms time window and in the 35-45 Hz (low gamma) frequency range in a 400-600 ms time window. Coherent (average of old-coherent and new-coherent) was compared to incoherent (average of old-incoherent and new-incoherent) in the 60-80 Hz (high gamma) frequency range in a 500-1000 ms time window. As the cluster-based permutation test is designed to compare two conditions at a time, we tested for an interaction effect by comparing the difference between coherent and incoherent in the old condition to the same difference in the new condition.

Non-preregistered analyses

TF analysis of new-neutral proper names

We performed exploratory time-frequency analysis on the conditions new-neutral, new-coherent and old-coherent in the same way as described in the section *EEG recording and preprocessing*. A cluster-based permutation test was used to compare new-neutral to both new-coherent and old-coherent in the 4-7 Hz (0-1000 ms), 35-45 Hz (400-600 ms) and 60-80 Hz (500-1000 ms) frequency ranges. Settings for the permutation test were identical to those described in the section *Statistical analysis of oscillatory power*.

TF analysis of ERP signal

To rule out the possibility that event-related potential activity contaminated our time-frequency analyses, we performed time-frequency analysis on the ERP signal that was obtained after within-subject averaging (see Wang et al., 2016 for a similar approach). As discussed in the introduction, ERPs contain phase-locked activity only, whereas time-frequency data contains both phase-locked and non-phase-locked activity. If the time-frequency results are driven by phase-locked activity, TF analysis of the ERP signal is expected to show a similar pattern as the TF analysis that was based on single-trial data. TF analysis of the subject-averaged ERP data was done in a similar way as described in the section *EEG recording and preprocessing*. A cluster-based permutation test was then used to compare the differences between old and new within the 4-7 Hz frequency range and a 300-500 ms time window.

Beamformer source localization

In an attempt to identify the sources underlying the observed differences in the 4-7 Hz theta and 60-80 Hz gamma oscillatory power, we applied a beamformer technique called Dynamical Imaging of Coherent Sources (Gross et al., 2001). This method uses a frequency-domain implementation of a spatial filter to estimate the source strength at a large number of previously computed grid locations in the brain. These grid locations are points in a three-dimensional grid that forms a discretized representation of a brain volume. Because the increase in 4-7 Hz theta activity for old compared to new names was most pronounced between 300-500 ms post-CW onset, this time period was subjected to source reconstruction. Following Nieuwland and Martin (2017), the increase in 60-80 Hz gamma activity for coherent compared to incoherent names was analyzed within a 500-1000 ms interval post CW-onset. The procedure and settings of the beamformer approach are adopted from Nieuwland and Martin (2017).

In addition to these condition-specific time windows, we extracted the data of all conditions in a 500-300 ms pre-CW baseline window. All data was re-referenced to the average of all electrodes. We identified the theta-activity to be most prominent between 4-7 Hz and therefore performed time-frequency analysis on 5 Hz, using a Hanning taper with ± 2 Hz spectral smoothing. In the gamma time window, we estimated power at 70 Hz, using a Slepian sequence taper with ± 10 Hz spectral smoothing.

We aligned the electrode positions of the montage to a standard Boundary Element Method (BEM) head model, which is a volume conduction model of the head based on an anatomical MRI (magnetic resonance imaging) template (Oostenveld, Praamsta, Stegeman & van Oosterom, 2001). This head model was subsequently discretized into a three-dimensional grid with a 5 mm resolution, and for each grid point an estimation of source power was calculated. For the 5 Hz and the 70 Hz frequencies-of-interest separately, a common inverse filter was computed on the basis of the combined dataset containing the pre-CW and post-CW intervals of both conditions (i.e., old-new for 5 Hz, coherent-incoherent for 70 Hz), which was then separately applied to all trials of each condition in order to estimate source power. After averaging over trials, we computed the difference between post-CW and pre-CW activity for each condition separately in the following way: (post-CW - pre-CW)/pre-CW. In order to visualize the estimated activity, we computed grand averages over subjects and subsequently interpolated the grid of the estimated power values to the anatomical MRI.

These estimates of source power were subjected to statistical analysis by means of a cluster-based permutation test (see section *Statistical analysis of oscillatory power*). On each source location in the three-dimensional grid we performed a one-sided dependent samples t-test (at $\alpha = .05$, yielding uncorrected p-values) on trial-averaged data of respectively old and new (5 Hz) and coherent and incoherent (70 Hz). Neighboring grid points with significant t-values were clustered. A cluster-level test statistic was calculated by summing the individual t-values within each cluster and evaluated relative to a permutation distribution that was based on 1000 randomizations of the same dataset. In order to localize the spatial coordinates of the areas exhibiting significant differences, we interpolated only the t-values of the significant, clustered source points to the anatomical MRI. We identified brain areas using a template atlas (Tzourio-Mazoyer et al., 2002).

Results

ERP analysis

At the N400 region-of-interest, the effect of coherence was similar in the old and new conditions, $\chi^2 = 0.97$, $p = .33$. In addition, coherent and incoherent names elicited a similar N400 (mean difference = 0.20, *SE* = 0.19), $\chi^2 = 1.16$, $p = .28$. New names elicited a more negative N400 than old names (mean difference = 2.14, *SE* = 0.19), $\chi^2 = 127.45$, $p < .001$.

At the LPC region-of-interest, the effect of coherence was similar in the old and new conditions, $\chi^2 = 3.01$, $p = .08$. In addition, coherent and incoherent names elicited a similar LPC (mean difference = 0.12, *SE* = 0.19), $\chi^2 = 0.41$, $p = .52$. New names elicited a more positive LPC than old names (mean difference = 0.73, *SE* = 0.19), $\chi^2 = 14.03$, $p < .001$. Figure 2 contains the ERPs and corresponding scalp distributions of the difference between old and new (2A) and coherent and incoherent (2B).

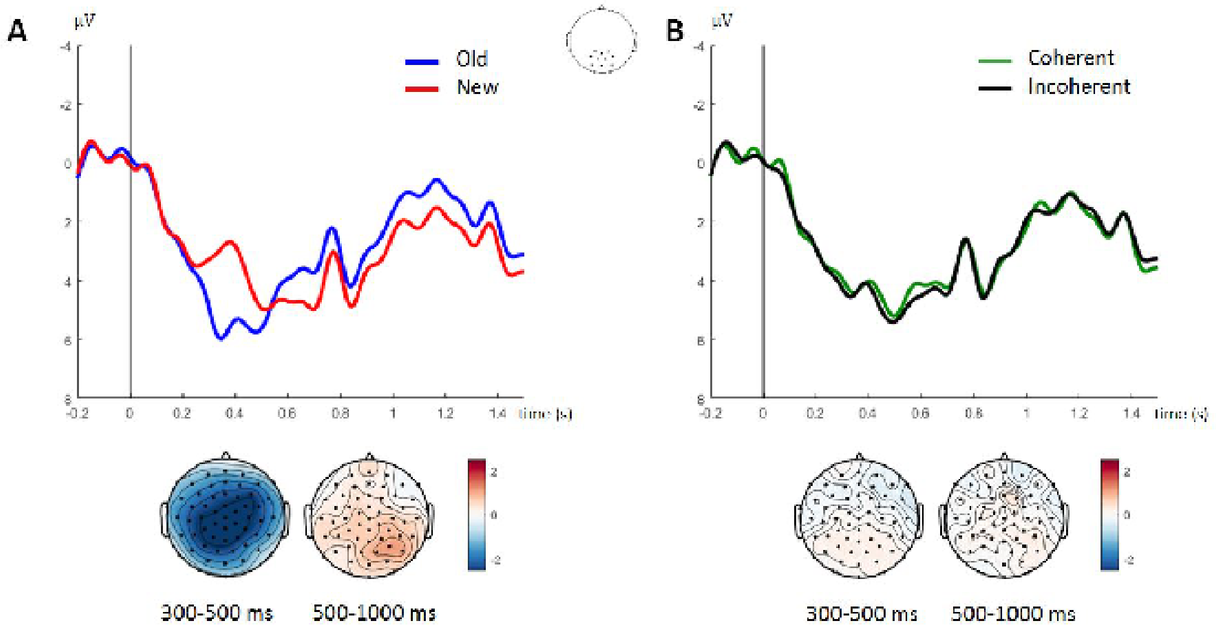

Fig. 2. N400 (300-500 ms) and LPC (500-1000 ms) responses, averaged over the centroparietal cluster, as a function of (A) givenness and (B) coherence. Scalp topographies represent the difference between (A) old and new and (B) coherent and incoherent. ERPs are low-pass filtered at 10 Hz for presentation purposes only.

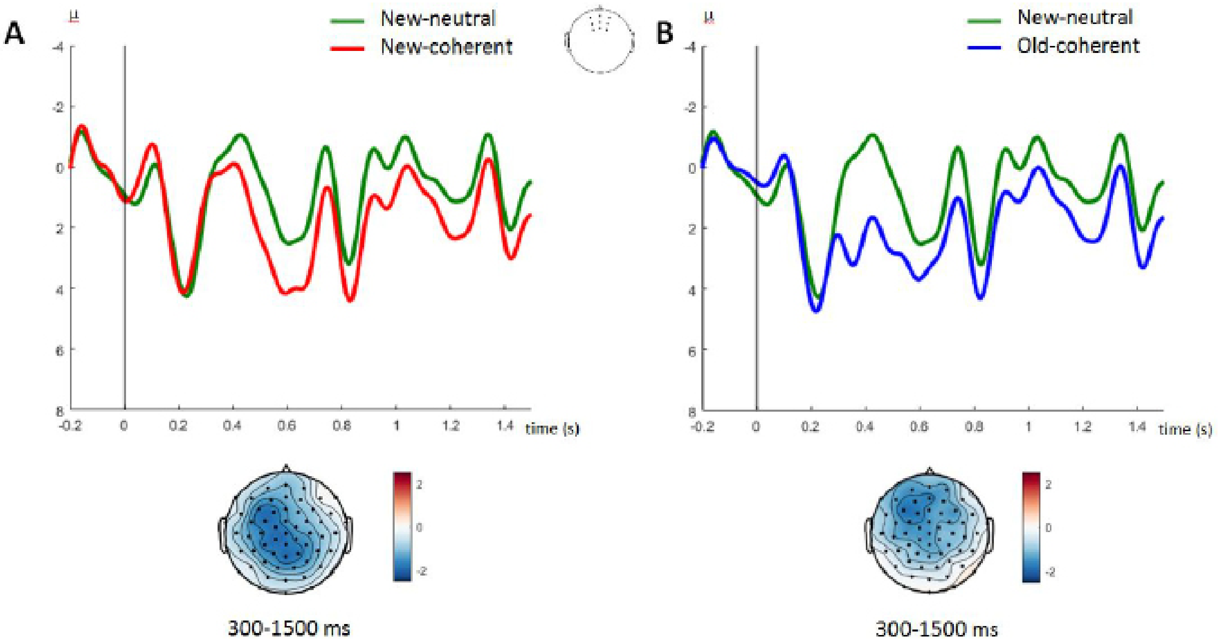

Fig. 3. Nref (300-1500 ms) responses, averaged over the frontal cluster, for the comparisons between new-neutral and new-coherent (A) and old-coherent (B). Scalp topographies reflect the difference between (A) new-neutral and new-coherent and (B) new-neutral and old-coherent. ERPs are low-pass filtered at 10 Hz for presentation purposes only.

At the Nref region-of-interest, the average ERP to new-neutral was significantly more negative than the average ERP to old-coherent (mean difference = 1.45, *SE* = 0.36), $\chi^2 = 14.61$, $p < .001$. Similarly, the average ERP to new-neutral was significantly more negative than the average ERP to new-coherent (mean difference = 1.17, *SE* = 0.34), $\chi^2 = 10.46$, $p = .001$. The ERPs and corresponding scalp distributions are shown in Figure 3.

TF analysis

In the 4-7 Hz (theta) frequency range, we found significantly larger theta-band power in the old compared to the new condition ($p = .034$). TF representations and scalp topography are presented

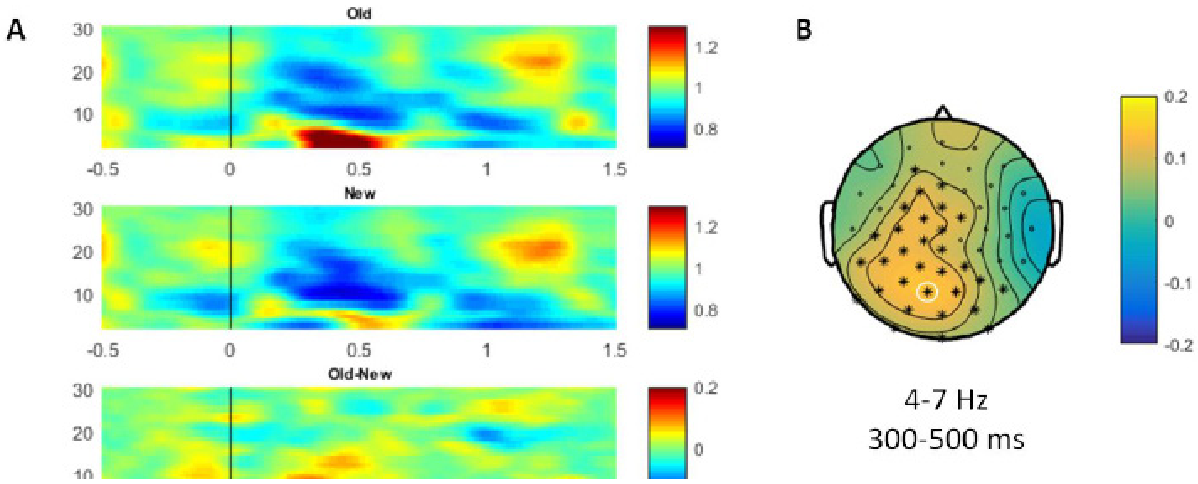

Fig. 4. Results of the TF analysis in the frequency range of 2-30 Hz. (A) TF representations (parietal-midline electrode 40) for old, new and the difference between old and new. (B) Topographical distribution of the 4-7 Hz theta effect in the 300-500 ms time window. Electrodes participating in the significant cluster are marked by an asterisk (*).

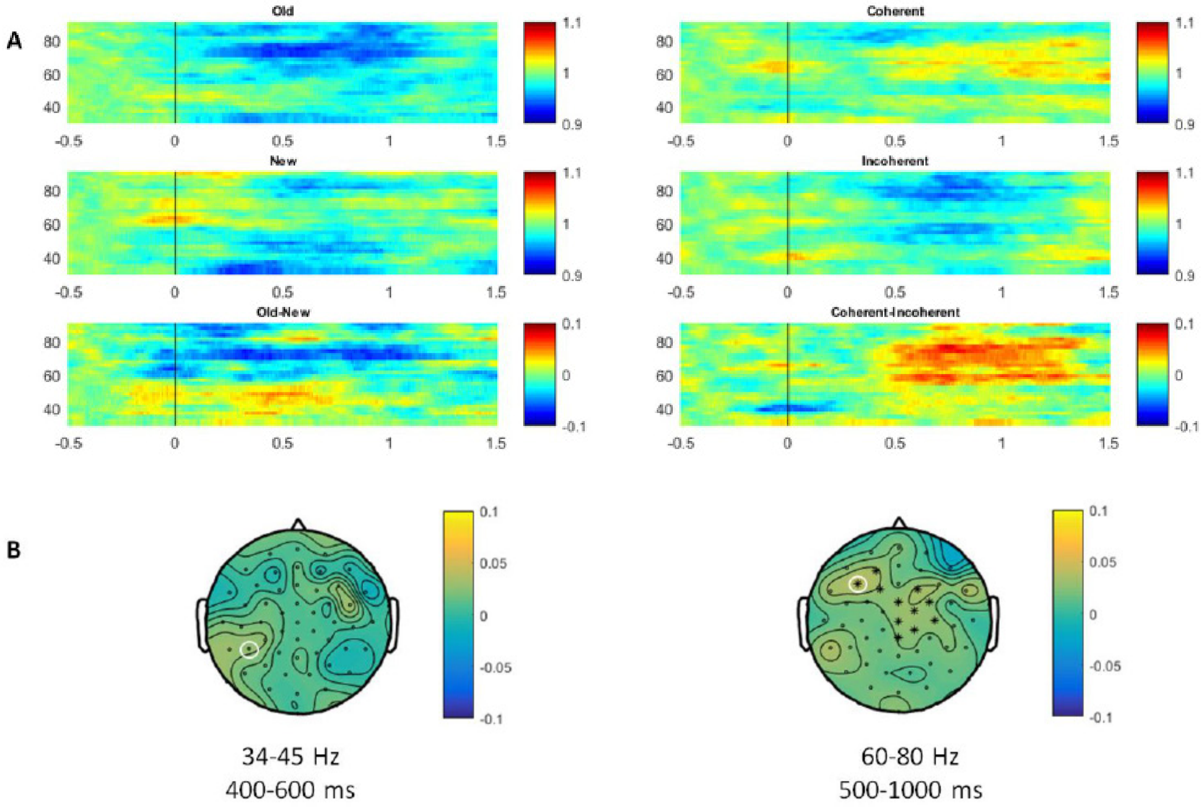

Fig. 5. Results of the TF analysis in the frequency range of 30-90 Hz. (A) The left panel shows the TF representations (left parietal electrode 42) of old, new and the difference between old and new. The right panel shows the TF representations (left frontal electrode 45) of coherent, incoherent and the difference between coherent and incoherent. (B) Topographical distributions for the 34-45 Hz old-new difference in the 400-600 ms time window (left) and for the 60-80 Hz coherent-incoherent difference in the 500-1000 ms time window (right). Electrodes participating in the significant cluster are marked by an asterisk (*).

in Figure 4.

In the 35-45 Hz (low gamma) frequency range, no significant clusters were observed for the contrast old-new (Figure 5a). In the 60-80 Hz (high gamma) frequency range, we observed significantly larger power in the coherent condition than in the incoherent condition ($p = .04$) within the preregistered time window of 500-1000 ms (Figure 5b). No significant clusters were observed for the interaction between givenness and coherence.

Exploratory analyses

TF comparisons of neutral and coherent conditions

In the 400-600 ms time window, old-coherent proper names elicited significantly stronger 35-45 Hz gamma-band synchronization than proper names in the new-neutral condition ($p = .004$). None of the other contrasts yielded significant differences. These results are visualized in Figure 6.

Beamformer source localization

A beamformer procedure was applied to localize the sources of the 4-7 Hz theta and 60-80 Hz gamma effects. We first applied a spatially unrestricted cluster-based permutation test to the power differences in the entire source space. This did not yield significant sources for the theta effect or the high gamma effect.

We therefore performed both literature-driven and data-driven exploratory region-of-interest analysis of the gamma effect (Figure 6A). Nieuwland and Martin (2017) localized the source of their high gamma activity to left frontal-temporal regions, encompassing inferior frontal lobe, inferior temporal lobe and anterior temporal lobe. We performed exploratory region-of-interest (ROI) analysis by restricting a cluster-based permutation test to these regions. It revealed a significant difference between coherent and incoherent, $p = .047$. Within this ROI, the difference was only significant in the left inferior frontal lobe (LIFG). However, as the effect seems to extend into dorsal regions of the frontal cortex

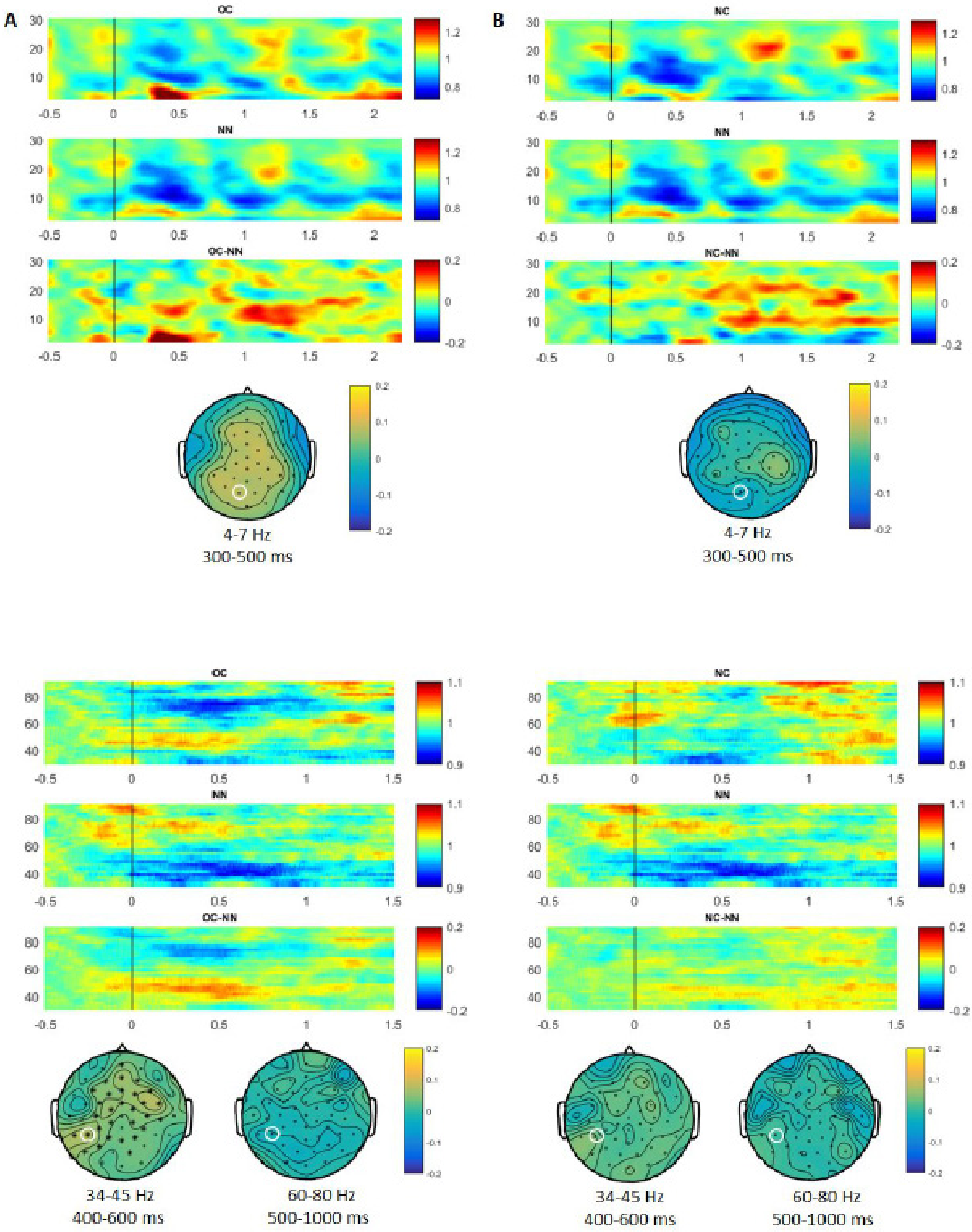

Fig. 6. Results of the TF analysis in the new-neutral (NN), old-coherent (OC) and new-coherent (NC) conditions. (A) TF representations of old-coherent, new-neutral and the difference between old-coherent and new-neutral, in both the low and high frequency ranges (parietal-midline electrode 40). (B) TF representations of new-coherent, new-neutral and the difference between new-coherent and new-neutral, in both the low and high frequency ranges (left parietal electrode 42). Below each TF representation are the topographical distributions of the respective differences. Electrodes participating in the significant cluster are marked by an asterisk (*).

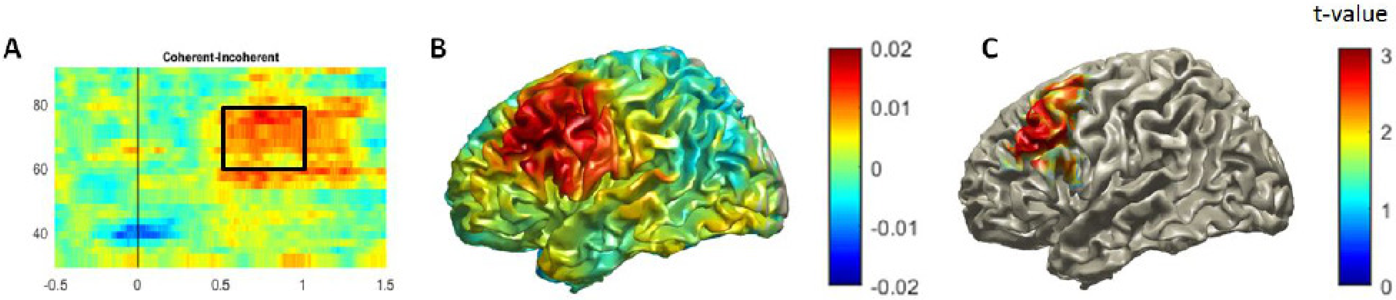

Fig. 7. Source localization results for the 60-80 Hz gamma effect. (A) TF representation of the difference between coherent and incoherent (black outline indicates the TF window of interest). (B) Surface plot of the power differences. (C) Surface plot of the ROI-based statistical results. Colors represent t-values, masked for significance.

(Figure 7B) and does not encompass any areas in the temporal lobe, we performed additional data-driven ROI analysis on the entire left frontal lobe in order to explore where the effect was strongest. Again, a significant difference between conditions was found (p = .027), in an area encompassing the LIFG and the left medial frontal gyrus. The source localization results are shown in Figure 7.

TF of ERP data

TF analysis of the ERP data revealed no significant differences between old and new (p > .1). Figure 8 presents the results of the TF analysis based on single-trial data and based on averaged ERP data. The TF representation of the ERP data seems to show a difference in power only up to 3 Hz, while we analyzed only frequencies within the frequency range of 4-7 Hz. Therefore, we believe that the theta old-new difference was not driven by phase-locked activity but rather represents 'true' oscillatory activity.

Discussion

In the present EEG study, we examined event-related potentials (ERPs) and neural oscillations in order to investigate the involvement of language and memory processes in anaphor comprehension. More specifically, we aimed to test the idea that gamma-band synchronization in response to coherent referential dependencies reflects the workings of the recognition memory and language networks. Subjects were presented two-sentence mini-discourses in which the interpretation of anaphoric (old/repeated) and non-anaphoric (new) proper names was either coherent or incoherent with respect to the preceding discourse. As expected, we observed that in comparison to new names, repeated names elicited an attenuated N400 followed by a reduced Late Positive Component (LPC). Surprisingly, the ERPs in response to discourse-coherent and discourse-incoherent names did not show any differences. Proper names in the new-neutral condition elicited an Nref effect in comparison to both coherent conditions. In the time-frequency signal, we

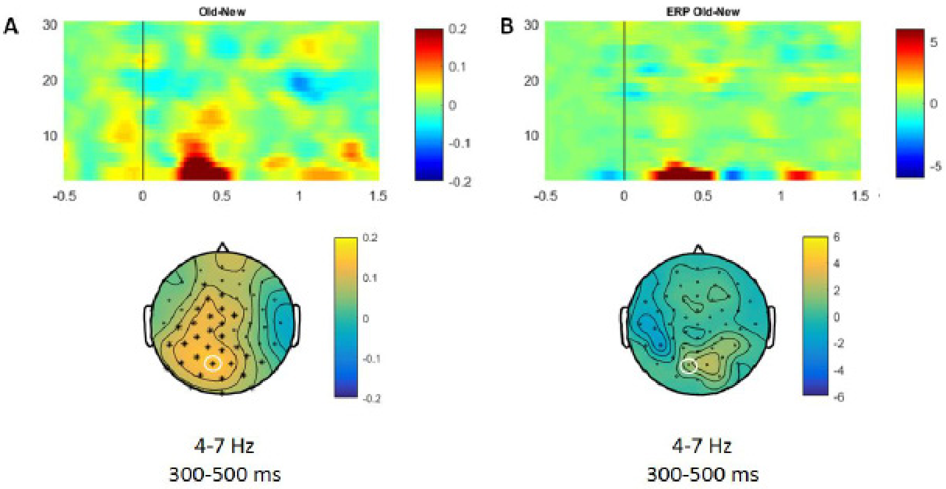

Fig. 8. Comparisons of TF representations (parietal-midline electrode 40) (A) on single trials and (B) on averaged ERP data. Topographical distributions of the 4-7 Hz average difference in the 300-500 ms time window are provided in the lower panel.

observed larger theta (4-7 Hz) power for repeated compared to new names. In the 60-80 Hz gamma range, we observed an increase in synchronization for coherent compared to incoherent proper names, which was most strongly associated with areas in the left frontal lobe. Exploratory analyses revealed an increase in 35-45 Hz gamma synchronization for old-coherent compared to new-neutral proper names.

Givenness and the N400-LPC complex

The expected biphasic ERP pattern observed for givenness is consistent with frameworks that view reference resolution as a two-stage process. These distinguish between a first stage at which the referent is lexically identified (i.e., reactivated from working memory), and a second stage at which the anaphor is integrated into the discourse model (Almor & Nair, 2007). In the current experiment, lexical identification of the referent of repeated proper names was facilitated because the antecedent had just been processed and was still available in working memory by the time the critical name was presented (see Wang & Yang, 2013 for a similar conclusion). The N400 effect in our experiment thus seems to be related to identifying the referent and reactivating its representation from working memory, a process that is facilitated for repeated names. New names do not reside in working memory and therefore elicit a large N400 component. The LPC might reflect updating of the discourse model by establishing an independent referential representation for the new name (Burkhardt, 2006, 2007; Schumacher, 2009; Schumacher & Hung, 2012; Wang & Schumacher, 2013; see also Kaan et al., 2007).

The Nref effect and dependency formation

Proper names in the new-neutral condition elicited a larger Nref than both the old-coherent and new-coherent proper names. Although the old-coherent names additionally elicited an attenuated N400 compared to new-neutral names, the fact that there was also an Nref for the comparison between new-neutral and new-coherent indicates that the difference between new-neutral and old-coherent is not solely due to the downstream consequences of name repetition. The Nref has been argued to reflect processes involved in resolving referential ambiguity (e.g., Nieuwland et al., 2007b; Van Berkum, 2009), but as the new-neutral names in our experiment are not genuinely ambiguous, we do not find this interpretation particularly compelling. Instead, we will argue that the Nref effects in response to our stimuli reflect difficulty in forming referential dependencies. Our interpretation derives from the cue-based retrieval framework (McElree, 2000, 2006; McElree et al., 2003), in which it is argued that the second element in a referential dependency (i.e., the anaphor) triggers reactivation of already encoded information that is held in working memory (i.e., the antecedent), which is addressable by virtue of content overlap between anaphor and antecedent (i.e., retrieval cues; Lewis & Vasishth, 2005; Lewis, Vasishth & Van Dyke, 2006). With respect to our stimuli in the new-neutral condition, we suggest that retrieval cues on the proper name trigger the reactivation of the reference group, which contains a sufficient amount of content overlap with the proper name (i.e., in terms of gender, animacy, etc.). However, because the match between anaphor and antecedent is not perfect, dependency formation does not run smoothly, as reflected in the Nref effect.

Evidence in favor of our interpretation of the Nref in terms of retrieval difficulty and its consequences for dependency formation comes from three recent ERP studies. First, Martin, Nieuwland and Carreiras (2012, 2014) investigated ERPs in response to elliptic determiners in Spanish, which either did or did not agree with their antecedent in terms of grammatical gender (e.g., … t-shirt$_{\text{FEM}}$ ... another$_{\text{MASC}}$). They additionally examined whether this process was modulated by the gender of a structurally unavailable local attractor noun. Both studies found that for fully grammatical sentences with unambiguous elliptic determiners the gender of the attractor modulated the amplitude of the Nref (albeit in opposite direction in both studies). These findings indicate that the gender of the attractor can interfere with ellipsis-based retrieval of the correct antecedent, even in the absence of referential ambiguity, and that the Nref might be an electrophysiological correlate of attempted retrieval and subsequent dependency formation. Similarly, Karimi, Swaab and Ferreira (2018) showed that retrieval difficulty, modulated as a function of the representational richness of antecedents (e.g., 'the actor' vs. 'the actor who was visibly upset') modulates the amplitude of the Nref. In all, this suggests that the Nref in our experiment represents difficulty establishing a referential dependency. The observation that such difficulty is also perceived in response to proper names provides evidence against Barkley et al.'s (2015) proposal that proper names do not trigger back association.

One caveat to our interpretation is that the ERPs in response to the new-coherent proper names did not show signs of referential dependency formation, although these conditions also contained a reference group (e.g., David and Peter are the worst players *in the football team*) to which the new proper name could have been linked. Two differences between the new-neutral and new-coherent condition might explain this result. That is, in the context sentence of the neutral condition, the reference group was the grammatical subject and it did not contain proper names. These differences might be relevant, because both factors have been shown to increase the discourse prominence of the denoted referent (Gordon & Hendrick, 1998; Gordon et al., 1999; Sanford, Moar & Garrod, 1988). As a result, the reference group in the new conditions might not have been accessible enough to be considered available for co-reference. We are currently designing a follow-up experiment in which the neutral condition is adjusted such that the presence of proper names is manipulated, and the reference group is mentioned more explicitly and thereby possibly made more available for co-reference. Note that this does not affect our interpretation of the Nref in the new-neutral condition in terms of dependency formation, but merely aims to answer the question why this process was not triggered in the new-coherent condition.

Absence of ERP effects for discourse coherence

To our surprise, we did not observe an effect of discourse coherence on the N400. A first possible interpretation of the absence of a coherence effect could be that participants did not notice the incoherence because they were not engaged enough. During the post-experiment debriefing all participants reported to have noticed the discourse incoherence, they all scored very high on the comprehension questions (average percentage correct of 92%), and the Nref observed in the comparisons between the neutral and coherent conditions indicates that participants were engaged in co-reference processes. This suggests that participants were adequately paying attention and did notice the incoherence, at least for some items. Yet, in order to check whether the coherence manipulation was effective for each individual item, we are currently setting up a norming test in which an additional group of participants will be asked to rate the coherence of each discourse item (e.g., on a 5-point Likert scale, as done by Wang, Verdonschot & Yang, 2016).

Two alternative, but related explanations are offered for the absence of a coherence effect. First, the very strong repetition effects (i.e., having a peak-to-peak difference of approximately 3 μV) might have washed out any effects of coherence. In terms of the absence of any difference between old-coherent and old-incoherent, it is conceivable that name repetition produced a ceiling effect, such that the addition of a coherent discourse did not further reduce the N400. This does not have any bearing on the absence of an N400 difference between new-coherent and new-incoherent, which, assuming that the N400 reflects semantic retrieval, had been hypothesized to elicit a similar N400 anyway (Kutas & Federmeier, 2000, 2011; Van Berkum, 2009). Also note that this inseparability in the time domain is perfectly consistent with the fact that we observed coherence-induced changes in the high gamma range, which are separable from the oscillatory effects of givenness in the time-frequency domain by virtue of different frequency characteristics. An alternative possibility is that the link between the initially meaningless proper names and the associated information was not strong enough to immediately affect processing difficulty. The relative coherence of each item hinged on the strength of the association between proper name and characteristic information described in the first sentence, which in turn is highly dependent on how semantically constraining the sentence is. Word-learning studies have shown that the meaning of novel words can be acquired very fast, but only when these novel words are learned in a strongly constraining context from which their meaning can easily be derived (Borovsky, Kutas & Elman, 2010; Mestress-Missé, Rodriguez-Fornells & Münte, 2007). On a similar note, Wang and Yang (2013) used a two-sentence discourse context to set up an explicit contrast between two newly introduced discourse entities. Linguistic contrast is known to facilitate word learning (Au & Markman, 1987; Kupferborg & Ohlstain, 1996) and is likely to have affected the fast mapping between proper name and characteristic information in the Wang and Yang (2013) study too. On top of that, participants in their study had to judge the congruence of the whole discourse after each trial, making it essentially inevitable that the incoherence was noticed. It is certainly possible that the proper name's meaning in our experiment was not yet established enough to yield immediate processing difficulty in the case of a discourse-incoherent interpretation. This, too, is compatible with the fact that we observed coherence effects in the time-frequency signal. Event-induced oscillations do not necessarily have to be time-locked in order to be picked up by time-frequency analysis. As long as the latency variability is not too large and does not exceed the length of the taper used for convolution with the EEG time course (Tallon-Baudrey & Bertrand, 2000), time-frequency analysis can pick up induced oscillations, albeit in relatively time-smoothed appearance (Cohen, 2014). The fact that the high gamma effects seem to be smoothed in time, extending beyond the expected time window of 1000 ms (denoted by the black contour in Figure 7A), is compatible with this possibility.

Increased theta-band synchronization for repeated names

Compared to new names, given names elicited a widespread increase in theta-band synchronization that was most prominent over left parietal electrodes. Our time-frequency analysis of the ERP signal in the N400 time window showed that the theta effect was not driven by phase-locked activity, instead reflecting ‘true’ oscillatory activity (Wang et al., 2012). This is in agreement with the findings that the given/new ERP and theta effects are independent phenomena (Jacobs et al., 2006; Klimesch et al., 2000).

As briefly discussed in the introduction, one line of research has related theta oscillations in language processing to retrieval from semantic long-term memory (Bastiaansen et al., 2002, 2005, 2008), where theta synchronization is increased when retrieval is difficult (Hagoort et al., 2004; Hald et al., 2006). As repeated and new proper names place similar demands on retrieval from long-term memory, this is not likely to have caused the theta effects. If anything, retrieval of new names would be more difficult than retrieval of repeated names, suggesting an increase in theta for new compared to repeated names, which is the opposite of what we found.

Rather, we interpret the theta effect as reflecting successful retrieval from working memory. To reiterate, studies of recognition memory have found an increase in theta-band synchronization for correctly remembered targets compared to correctly rejected distractors (Burgess & Gruzelier, 1997, 2000; Chen & Caplan, 2016; Jacobs et al., 2006; Klimesch et al., 1997, 2000, 2006; Osipova et al., 2006; Van Strien, 2005, 2007), suggesting that theta oscillations might reflect a relational process that matches the probe to a representation held in working memory (Chen & Caplan, 2016; Jacobs et al., 2006). This is in line with the idea that theta oscillations also underlie retrieval of information in during online language processing (Covington & Duff, 2016; Duff & Brown-Schmidt, 2012; Meyer et al., 2015).

No influence of givenness on low gamma oscillations

Against our prediction, we did not observe a difference between repeated and new names in the 34-45 Hz (low) gamma range. Interestingly, however, exploratory analyses revealed that old-coherent proper names elicited larger low gamma synchronization than proper names in the new-neutral condition. It should be noted that our prediction was solely based on Nieuwland and Martin’s interpretation of their gamma effects for coherent anaphors. They localized the origin of these effects to the left posterior parietal cortex (LPPC), an area that has been related to successful memory retrieval (e.g., Öztekin et al., 2008; Wagner et al., 2005) and the ability to make old/new judgments (Gonzalez et al., 2015). However, activity in this area is often related to aspects of memory that are arguably different from the memory mechanisms that underlie language comprehension. For instance, fMRI (functional magnetic resonance imaging) studies have indicated that activity in the LPPC relates to a subjective perception of memory strength (Hutchinson, Uncapher & Wagner, 2015) and the phenomenological experience of remembering (Wagner et al., 2005). In addition, the LPPC has been implicated in recovery of information about the temporal ordering of to-be-remembered items, which is thought to require a slow serial search operations (Öztekin et al., 2008). As argued before, there is good reason to believe that the memory mechanisms underlying language comprehension work via direct access rather than serial search (McElree, 2000, 2006; McElree et al., 2003). Combined with the absence of an old/new effect in the low gamma range, this suggests that language comprehension does not rely on the memory-preserving functions of the LPPC. This begs the question what the gamma effects do reflect. One commonality between our findings and those by Nieuwland and Martin is that both were elicited by a comparison that also elicited an Nref effect. Although the specific processes underlying both Nref effects are probably different (i.e., dependency formation vs. resolving referential ambiguity), this might indicate that the gamma effects are related to general processes involved in resolving a (linguistically) complicated situation. Further research is needed to find out whether the low gamma effects are specific to referential processes or whether they reflect domain-general cognitive mechanisms involved in complex tasks, as has been proposed for the LPPC (e.g., Chein, Ravizza & Fiez, 2003).

Increased high gamma synchronization for coherent names

In line with our expectations, we observed an increase in 60-80 Hz gamma-band synchronization in response to discourse-coherent compared to discourse-incoherent proper names. Literature-based exploratory ROI analysis revealed that the effect was generated by left frontal regions, encompassing the left inferior frontal gyrus (LIFG). Previous studies of semantic unification on discourse-level have shown that the language processor immediately uses information from both sentence-level and discourse-level sources (Van Berkum et al., 1999a, 2003) and integrates these in a 'single unification space', predominantly active in the LIFG (Hagoort & Van Berkum, 2007; Hagoort & Indefrey, 2014). Semantic unification on the sentence-level has been shown to modulate oscillations in the gamma band, whereby it is generally observed that synchronization is increased whenever semantic unification is successful (Bastiaansen & Hagoort, 2015; Hald et al., 2006; Peña & Melloni, 2012; Penolazzi, Angrili & Job, 2009). We provide converging evidence that combines these patterns: discourse-level manipulations of semantic unification modulate gamma oscillations in the LIFG.

It has recently been noted that these gamma-band modulations might be more easily explained in terms of prediction rather than semantic unification (Lewis & Bastiaansen, 2015; Lewis, Wang & Bastiaansen, 2015; Wang, Zhu & Bastiaansen, 2012). Bastiaansen and colleagues propose that a match between pre-activated representations and incoming language material translates into gamma-band synchronization. Specific to our stimuli, it could be argued that the discourse-coherent proper names were predictable and therefore elicited a gamma-band increase. However, new proper names were never predictable, whether coherent with the preceding discourse or not, suggesting that some of our conditions do not lend themselves for predictive processes. In addition, we localized the source of the gamma effect to the LIFG, which has been strongly linked to unification operations (Hagoort, 2005; Hagoort & Indefrey, 2015), and differentiates between discourse-coherent and incoherent anaphors (Hammer, Jansma, Tempelmann & Münte, 2011). Context-based predictive processes, in contrast, typically show up as activation in medial temporal regions (e.g., Lau & Nguyen, 2015). Last, no effect of coherence was found on the N400, which has been strongly linked to predictability (Kutas & Hillyard, 1984). Although one could argue that the absence of any coherence-related differences on the N400 also constitutes evidence against the view that our gamma findings reflect semantic unification, in the preceding paragraphs we suggested that the absence of coherence-related modulations of the N400 might have been related to time variability in the moment at which the incoherence was noticed. If preparatory processes had been able to preactivate the (old-)coherent proper names, one would expect the incoherence to be noticed as soon as the word had been recognized (i.e., predicted representations might be given a 'head-start' in activation; Lau & Nguyen, 2015), and not leading to a possible incoherence response that has a variable latency.

Thus, while previous studies have shown the contextually constraining effects of discourse on semantic unification in the time domain (i.e., modulations of the N400 amplitude; e.g., Van Berkum et al., 1999a, 2003), the present study is the first to demonstrate these effects of discourse coherence in the time-frequency domain.

Conclusion

The present EEG study used ERPs and neural oscillations to study the involvement of recognition memory and semantic unification in anaphor comprehension via respectively givenness and coherence. We manipulated givenness by utilizing the ability of proper names to introduce new reference and maintain old reference, and found that it modulates oscillatory activity in the theta band. Coherence of proper names was reflected in gamma-band synchronization. In all, our study is the first to show that givenness and coherence of discourse-level anaphors modulates oscillatory synchronization in separated frequency bands, thereby encouraging future time-frequency research into the role of memory mechanisms in discourse-level language comprehension.

References

Almor, A., & Nair, V. A. (2007). The form of referential expressions in discourse. *Language and Linguistics Compass, 1*(1-2), 84-99.

Almor, A., Nair, V. A., Boiteau, T. W., & Vendemia, J. M. C. (2017). The N400 in processing repeated name and pronoun anaphors in sentences and discourse. *Brain and Language, 173*, 52-66.

Au, T. K., & Markman, E. M. (1987). Acquiring word meanings via linguistic contrast. *Cognitive Development, 2*, 217-236.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Barkley, C., Kluender, R., Kutas, M. (2015). Referential processing in the human brain: an event-related potential (ERP) study. *Brain Research, 1629*, 143-159.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bastiaansen, M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in brain research*, *159*, 179-196.

Bastiaansen, 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.

Bastiaansen, M. C., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: Oscillatory neuronal dynamics. In S. J. Luck, & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 31-50). New York, NY: Oxford University Press.

Bastiaansen, M., Oostenveld, R., Jensen, O., & Hagoort, P. (2008). I see what you mean: Theta power increases are involved in the retrieval of lexical semantic information. *Brain and Language, 106*, 15–28.

Bastiaansen, M., Van der Linden, M., ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta responses are involved in lexicosemantic retrieval during language processing. *Journal of Cognitive Neuroscience, 17*, 530–541.

Bastiaansen, M., Berkum, J. J. A., & Hagoort, P. (2002). Event-related theta power increases in the human EEG during online sentence processing. *Neuroscience Letters, 323*, 13-16.

Bates, D. M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes.

Bertrand, O., & Tallon-Baudry, C. (2000). Oscillatory gamma activity in humans: a possible role for object representation. *International Journal of Psychophysiology*, *38*(3), 211-223.

Borovsky, A., Kutas, M., & Elman (2010). Learning to use words: event-related potentials index single-shot contextual word learning. *Cognition, 116*, 289-296.

Burgess, A.P., & Gruzelier, J.H., (1997). Short duration synchronization of human theta rhythm during recognition memory. *NeuroReport, 8*, 1039–1042.

Burgess, A.P., Gruzelier, J.H. (2000). Short duration power changes in the EEG during recognition memory for words and faces. *Psychophysiology, 37*, 596–606.

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

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport, 18*, 1851–1854.

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007a). Processing new and repeated names: effects of coreference on repetition priming with speech and fast rsvp. *Brain Research, 1146*, 172-184.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007b). The interplay of discourse congruence and lexical association during sentence processing: evidence from ERPs and eye tracking. *Journal of Memory and Language, 56*, 103-128.

Chen, Y. Y., & Caplan, J. B. (2016). Rhythmic activity and individual variability in recognition memory: theta oscillations correlate with performance whereas alpha oscillations correlate with ERPs. *Journal of Cognitive Neuroscience, 29*(1), 183–202.

Chein, J. M., Ravizza, J. A., & Fiez, J. A. (2003). Using neuroimaging to evaluate models of working memory and their implications for language processing. *Journal of Neurolinguistics, 16*, 315-339.

Cohen, M. X. (2014). *Analyzing neural time series data*. Cambridge, MA: MIT Press.

Covington, N. V., & Duff, M. C. (2016). Expanding the language network: direct contributions from the hippocampus. *Trends in Cognitive Sciences, 20*(12), 869-870.

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.

Duff, M. C., & Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Frontiers in human neuroscience*, *6*, 69.

Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: evidence from the N400. *Psychophysiology, 45*, 554-558.

Gonzalez, A., Hutchinson, J. B., Uncapher, M. R., Chen, J., LaRocque, K. F., Foster, B. L. ..., & Wagner, A. D. (2015). Electrocorticography reveals the temporal dynamics of posterior parietal cortical activity during recognition memory decisions. *Proceedings of the National Academy of Sciences, 112*(34), 11067-11071.

Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science, 22*(4), 389-424.

Gordon, P. C., Hendrick, R., Ledoux, K., & Yang, C. L. (1999). Processing of reference and the structure of language: an analysis of complex noun phrases.

Language and cognitive processes, 14, 353-379.

Gross, J., Kujala, J., Hämäläinen, 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, 98*(2), 694–699.

Hagoort, P. (2005). On Broca, brain, and binding. *Trends in Cognitive Science, 9*, 416-423.

Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, *37*, 347-362.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences, 4th ed.* (pp. 819-836). Cambridge, MA: MIT Press.

Hagoort, P., & Van Berkum, J. J. A. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B, 362,* 801-811.

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.

Hald, L. A., Bastiaansen, M. C., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, 96, 90–105.

Hammer, A., Jansma, B. M., Tempelman, C., & Münte, T. F. (2011). Neural mechanisms of anaphoric reference as revealed by fMRI. *Frontiers in Psychology, 2*(32).

Hutchinson, J. B., Uncapher, M. R., & Wagner, A. D. (2015). Increased functional connectivity between dorsal posterior parietal and ventral occipitotemporal cortex during uncertain memory decisions. *Neurobiology of learning and memory*, *117*, 71-83.

Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision making. *NeuroImage, 32*, 978-987.

Kaan, E., Dallas, A. C., & Barkley, C. M. (2007). Processing bare quantifiers in discourse. *Brain Research, 1146*, 199-209.

Karimi, H., Swaab, T., & Ferreira, F. (2018). Electrophysiological evidence for an independent effect of memory retrieval on referential processing. *Journal of Memory and Language, 102*, 68-82.

Kielar, A., Panamsky, L., Links, K. A., & Meltzer, J. A. (2015). Localization of electrophysiological responses to semantic and syntactic anomalies in language comprehension with MEG. *NeuroImage, 105*, 507-524.

Klimesch, W., Doppelmayr, M., Schimke, H., & Ripper, B. (1997). Theta synchronization and alpha desynchronization in a memory task. *Psychophysiology, 34*, 169–176.

Klimesch, W., Doppelmayr, M., Schwaiger, J., Winkler, T., & Gruber, W. (2000). Theta oscillations and the ERP old/new effect: independent phenomena? *Clinical Neurophysiology, 111*, 781–793.

Kupferborg, I., & Olshtain, E. (1996). Explicit contrastive instruction facilitates the acquisition of difficult L2 forms. *Language Awareness, 5*(3/4), 149-165.

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

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 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*, 621-647.

Lau, E. F., & Nguyen, E. (2015). The role of temporal predictability in semantic expectation: An MEG investigation. *Cortex*, *68*, 8-19.

Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (2007). Coreference and lexical repetition: neural mechanisms of discourse integration. *Memory & Cognition, 35*, 801-815.

Lewis, R. L., & Vasishth, S. (2005). An activation‐based model of sentence processing as skilled memory retrieval. *Cognitive science*, *29*(3), 375-419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in cognitive sciences*, *10*(10), 447-454.

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, *68*, 155-168.

Lewis, A. G., Wang, L., & Bastiaansen, M. (2015). Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction? *Brain and Language*, *148*, 51-63.

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology, 49*(4), 549-65.

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

Martin, A. E. (2016). Language processing as cue integration: grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology, 7*(120).

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *Neuroimage, 59*, 1859-1869.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2014). Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain and Language, 135*, 42-51.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*, 111-123.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Eds.), *The psychology of learning and motivation: Vol.*

46. The psychology of learning and motivation: Advances in research and theory (pp. 155-200). San Diego, CA, US: Elsevier Academic Press.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*(1), 67-91.

Mestres-Missé, A., Rodriguez-Fornells, A., Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex, 17*, 1858-1866.

Meyer, L. (2017). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *European Journal of Neuroscience.*

Meyer, L., Grigutsch, M., Schmuck, N., Gaston, P., & Friederici, A. (2015). Frontal–posterior theta oscillations reflect memory retrieval during sentence comprehension. *Cortex, 71*, 205–218.

Mitra, P. P., & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical journal*, *76*(2), 691-708.

Nieuwland, M. S. (2014). "Who's he?" Event-related brain potentials and unbound pronouns. *Journal of Memory and Language, 76*, 1–28.

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. (2007a). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience, 19*, 228–236.

Nieuwland, M. S., Petersson, K. M., & Van Berkum, J. J. A. (2007b). On sense and reference: examining the functional neuroanatomy of referential processing. *NeuroImage, 37*, 993-1004.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006a). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18*(7), 1098-1111.

Nieuwland, M. S., & Van Berkum, J. J. (2006b). Individual differences and contextual bias in pronounresolution: Evidence from ERPs. *Brain Research, 1118*, 155–167.

Nyhuis, E., Curran, T., 2010. Functional role of gamma and theta oscillations in episodic memory. *Neuroscience and Biobehavioral Reviews, 34*, 1023–1035.

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*, 2011, Article 156869.

Oostenveld, R., Praamstra, P., Stegeman, D. F., & van Oosterom, A. (2001). Overlap of attention and movement-related activity in lateralized event-related brain potentials. *Clinical Neurophysiology, 112*, 477-484.

Osipova, D., Takashima, A., Oostenveld, R., Fernàndez, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience, 26*, 7523-7531.

Öztekin, I., McElree, B., Staresina, B. P., Davachi, L. (2008). Working memory retrieval: contributions to the left prefrontal cortex, the left posterior parietal cortex, and the hippocampus. *Journal of Cognitive Neuroscience, 21*(3), 581-593.

Peña, M., & Melloni, L. (2012). Brain oscillations during spoken sentence processing. *Journal of Cognitive Neuroscience*, *24*(5), 1149-1164.

Penolazzi, B., Angrili, A., & Job, R. (2009). Gamma EEG activity induced by semantic violation during sentence reading. *Neuroscience Letters, 465*, 74-78.

Piai, V., Anderson, K. L., Lin, J. J., Dewar, C., Parvizi, J., Dronkers, N. F., & Knight, R. T. (2016). Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proceedings of the National Academy of Sciences*, *113*(40), 11366-11371.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, *25*(5), 762-776.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology, 22*, 642-647.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition of high- and low-frequency words. *Memory & Cognition, 18*, 367-379.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Science, 11*, 251-257.

Salmon, N., & Pratt, H. (2002). A comparison of sentence- and discourse-level semantic processing: an ERP study. *Brain and Language, 83*, 367-383.

Sanford, A. J., Moar, J., & Garrod, S. C. (1988). Proper names as controllers of discourse focus. *Language and Speech, 31*(1), 53-56.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*(1), 1-17.

Schumacher, P. B. (2009). Definiteness marking shows late effects during discourse processing: evidence from ERPs. *Lecture Notes in Artificial Intelligence, 5847*, 91–106.

Schumacher, P. B., & Hung, Y.-C. (2012). Positional influences on information packaging: insights from topological fields in German. *Journal of Memory and Language, 67*, 295–310.

St. George, M., Mannes, S., & Hoffman, J. E. (1994). Global expectancy and language comprehension. *Journal of Cognitive Neuroscience, 6*(1), 70-83.

Streb, J., Hennighausen, E., Rösler, F. (2004). Different anaphoric expressions are investigated by event-related potentials. *Journal of Psycholinguistic Research, 33*(3), 175-201.

Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience, 16*, 715-726.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … & Joliot, M.

(2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15*, 273–289.

Van Berkum, J. J. A. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things? In M. Carreiras, & C. Cliftin, Jr. (Eds.). *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 229-270). New York: Psychology Press.

Van Berkum, J. J. A. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276-316). Basingstoke: Palgrave Macmillan.

Van Berkum, J. J. A. (2012). The electrophysiology of discourse and conversation. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), The Cambridge handbook of psycholinguistics (pp. 589-614). New York: Cambridge University Press.

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999a). Semantic integration in sentences and discourse: evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657-671.

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

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

Van Berkum, J. J., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158-171.

Van Berkum, J. J. A., Zwitserlood, P., Bastiaansen, M. C. M., Brown, C. M., & Hagoort, P. (2004). So who's "he" anyway? Differential ERP and ERSP effects of referential success, ambiguity and failure during spoken language comprehension. *Supplement to the Journal of Cognitive Neuroscience, 16*, 70.

Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*, 701-718.

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractioning the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience, 3*(2), 131-150.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190.

Van Strien, J.W., Hagenbeek, R.E., Stam, C.J., Rombouts, S.A.R.B., & Barkhof, F. (2005). Changes in brain electrical activity during extended continuous word recognition. *NeuroImage, 26*, 952-959.

Van Strien, J. W., Verkoeijen, P. P. J. L., Van der Meer, N., & Franken, I. H. A. (2007). Electrophysiological correlates of word repetition spacing: ERP and induced band power old/new effects with massed and spaced repetitions. *International Journal of Psychophysiology, 66*, 205-214.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in cognitive sciences*, *9*(9), 445-453.

Wang, L., & Schumacher, P. B. (2013). New is not always costly: evidence from online processing of topic and contrast in Japanese. *Frontiers in Psychology, 4*(363).

Wang, L., Verdonschot, R., & Yang, Y. (2016). The processing difference between person names and common nouns in sentence contexts: an ERP study. *Psychological Research, 80*, 94-108.

Wang, L., & Yang, Y. (2013). Integrating the meaning of person names into discourse context: an event-related potential study. *PLOS ONE, 8*(12), e83206.

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.