

Activating words beyond the unfolding sentence: Contributions of event simulation and word associations to discourse reading

Florian Hintz^{a,*}, Antje S. Meyer^{a,b}, Falk Huettig^{a,c}

^a Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

^b Donders Institute for Brain, Cognition, and Behavior, Radboud University, Nijmegen, the Netherlands

^c Centre for Language Studies, Radboud University, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:

Discourse reading
ERPs
Associations
Event simulation

ABSTRACT

Previous studies have shown that during comprehension readers activate words beyond the unfolding sentence. An open question concerns the mechanisms underlying this behavior. One proposal is that readers mentally simulate the described event and activate related words that might be referred to as the discourse further unfolds. Another proposal is that activation between words spreads in an automatic, associative fashion. The empirical support for these proposals is mixed. Therefore, theoretical accounts differ with regard to how much weight they place on the contributions of these sources to sentence comprehension. In the present study, we attempted to assess the contributions of event simulation and lexical associations to discourse reading, using event-related brain potentials (ERPs). Participants read target words, which were preceded by associatively related words either appearing in a coherent discourse event (Experiment 1) or in sentences that did not form a coherent discourse event (Experiment 2). Contextually unexpected target words that were associatively related to the described events elicited a reduced N400 amplitude compared to contextually unexpected target words that were unrelated to the events (Experiment 1). In Experiment 2, a similar but reduced effect was observed. These findings support the notion that during discourse reading event simulation and simple word associations jointly contribute to language comprehension by activating words that are beyond contextually congruent sentence continuations.

1. Introduction

Written text comprehension is fast and efficient. Ziefle (1998) and Noyes and Garland (2008) estimated that the average adult reads prose at a speed of 250–300 words per minute (see Brysbaert, 2019, for a lower estimate). Why is it that we read text with such ease and at such a high speed? One reason, which has turned into a prominent feature of many models of sentence comprehension, is that we often pre-activate words beyond the unfolding sentence (e.g., Ferreira and Chantavarin, 2018; Pickering and Garrod, 2013; Pickering and Gambi, 2018; Schwanenflugel and Shoben, 1985; Van Berkum et al., 2005; Wicha et al., 2003, 2004).

Previous studies have led to a good understanding of the contents language users may pre-activate during comprehension (i.e. what is pre-activated, Arai and Keller, 2013; Chen et al., 2005; Federmeier et al., 2002; Laszlo et al., 2012; Rommers et al., 2013; Staub and Clifton, 2006) as well as the cues used to pre-activate linguistic or non-linguistic

information (e.g., Federmeier and Kutas, 1999; Knoeferle et al., 2005; Van Berkum et al., 2005). However, the underlying mechanisms, i.e. the processes connecting cues and contents, still remain poorly understood. In particular, experimental studies providing evidence for specific mechanisms contributing to (predictive) language comprehension are sparse. The goal of the present study was to map out the contributions of two proposed mechanisms, event simulation and simple associations, to discourse reading.

1.1. Event simulation

Researchers have long been highlighting the importance of event knowledge—referring to knowledge about real world events and regularities, as well as objects, persons and locations involved in them—for language and cognitive processing more broadly (Barsalou, 2008; Kahneman and Tversky, 1973; Mandler, 1984; Minsky, 1974; Rumelhart, 1980; Schank and Abelson, 1977; Zacks and Tversky, 2001). Concerning

* Corresponding author. Max Planck Institute for Psycholinguistics, P.O. Box 310 6500, AH, Nijmegen, the Netherlands.

E-mail address: florian.hintz@mpi.nl (F. Hintz).

the involvement of event knowledge in language processing, it has been proposed that language users engage in simulating real-world events and pre-activate linguistic representations that are likely to occur in that event (Elman, 2009; Elman and McRae, 2019; Huettig, 2015).

In an electroencephalography (EEG) study, Metusalem et al. (2012) reported experimental evidence demonstrating that event simulation contributes to activating referents that are likely to be mentioned in an unfolding discourse. They recorded the N400, which is a negative-going deflection in the EEG signal that peaks around 400 ms after stimulus onset. It is considered an index of semantic processing (Kutas and Hillyard, 1980; Kutas and Federmeier, 2011, for review). The N400 ERP component is typically distributed over centro-parietal electrodes. Participants in the study by Metusalem et al. read passages consisting of two introductory sentences and one target sentence. Taken together, the three sentences described typical events such as ‘The parents were very excited about their new baby girl. One of the first things they did was to get her baptized in their church. The baby liked baths, so she smiled when she was sprinkled with *water/priest/dentist* on her forehead’. The authors observed a three-way split in N400 amplitude: Expected targets (‘water’ in the example) elicited the smallest N400 amplitude. Contextually unexpected words (‘dentist’) elicited the largest N400 amplitude. Interestingly, contextually unexpected targets (‘priest’), which were related to the described event, elicited an attenuated N400 amplitude, which lay in between these two extremes. In their second experiment, participants read the target sentences presented in isolation (without the first two introductory sentences). Now the N400 elicited by the unexpected but event-related word (‘priest’) condition did not differ from the N400 elicited by the unexpected event-unrelated word (‘dentist’). Metusalem et al. concluded that readers used generalized knowledge to simulate the described event, thereby constraining the activation of potentially upcoming referents, i.e. event-related words beyond the most expected continuation of the target sentences. Specifically, they reasoned that the N400 reduction “[...] suggests that event knowledge also is an important knowledge source for driving linguistic prediction” (p. 559).

Using latent semantic analysis norms (LSA, Landauer and Dumais, 1997), Metusalem et al. conducted a post-hoc analysis assessing the potential influence of word associations on the observed pattern of results. To that end, they calculated the cosine distance between the target words and the entire preceding context. Event-related targets were significantly more strongly associated with the preceding words than the completely unrelated targets. In a next step, the authors removed twelve items, which showed the largest differences between LSA association scores for the event-related and event-unrelated targets, and re-ran their ERP analysis. The analysis revealed the same three-way split in N400 amplitude as previously, i.e. a significant difference between event-related and event-unrelated unexpected target words. Metusalem and colleagues thus concluded that the effect seen in the event-related condition was unlikely to be due to associative priming, i.e. simple word associations between words in the preceding context and the critical targets.

Importantly, while event simulation is likely to be a strong contributor to language comprehension, to be able to exclude the contribution of associations to discourse reading, a more direct test than a post-hoc analysis is needed. Observing evidence for the involvement of associations in discourse reading would be in line with theoretical frameworks that assume that simple word associations play a role in anticipatory language comprehension, albeit as one of multiple contributing mechanisms (Huettig, 2015; Kuperberg, 2007; Pickering and Gambi, 2018; Pickering and Garrod, 2013; cf. Bar, 2007, 2009).

1.2. Word associations

As described above, one way of operationalizing word associations, or the associative strength between two words, is by calculating their co-occurrence frequency based on large corpora of written texts (such as

LSA, Landauer and Dumais, 1997). Another common way is to use experimental tasks such as free association (De Deyne et al., 2013), where participants are provided with a cue word and have to generate one or multiple associates that come to mind. The general assumption is that words that co-occur frequently have a strong associative link and that such links are utilized in various cognitive tasks. Note that even though the nature of such links is often semantic, an associative relationship between two words does not necessarily require a semantic relation. For example, the classic doctor-nurse pair may be associatively related because of a semantic relationship and lexical co-occurrence, whereas pairs such as ‘good morning’ are associatively related purely on the basis of lexical occurrence. The present study concerns relationships of the former type, and we do not differentiate between purely semantic and purely associative (without a semantic relationship) links.

Neely (1991) proposed that the activation of one word leads to the retrieval/activation of another word in an automatic or ‘expectancy-based’ fashion. In a similar vein, Pickering and Garrod (2013), Huettig (2015), Kuperberg (2007), Pickering and Gambi (2018) assume that associations contribute to language comprehension, and specifically anticipatory processing. For example, these accounts assume that ‘church’ and ‘priest’ (as in the stimulus example by Metusalem et al., 2012) are words that often occur together; they thus have a strong associative link. The proposal is that activation spreading via the associative link, upon encountering one word, contributes to the (partial) pre-activation of the other word, irrespective of the unfolding sentence.

Electrophysiological studies investigating associative priming in word pairs such as ‘church-priest’ have demonstrated that target words elicit a reduced N400 amplitude when preceded by associated primes as compared to unassociated primes (Bentin et al., 1985; Bentin, 1987; Rugg, 1987). For example, Van Petten (2014) showed that free association strength (e.g. Nelson et al., 2004), and corpus-based measures of association, such as LSA, accounted for substantial amounts of variance in the amplitude of the N400 component elicited by the second word in 303 word pairs. These and many related behavioral studies, employing different paradigms and techniques (see Hutchinson, 2003; Neely, 1991, for review), suggest that associative relationships facilitate word processing.

A recurring issue in the current literature concerns the contribution of simple associations to comprehension and prediction of linguistic stimuli more complex than word pairs (see Ledoux et al., 2006, for review). To investigate this issue, previous studies have typically manipulated the associative relationship in word pairs (associated vs. unassociated) and the words’ fit with the sentential context (congruent vs. incongruent). Pioneering work was carried out by Van Petten (1993), who compared the influence of associations with the influence of sentential contexts on the amplitude of the N400 component. An example of her stimuli is given in (1). Associated words are marked in italics.

(1) **Associated/congruent:** When the *moon* is full it is hard to see many *stars* or the Milky Way.

Unassociated/congruent: When the *insurance* investigators found that he’d been drinking they *refused* to pay the claim.

Associated/anomalous: When the *moon* is rusted it is available to buy many *stars* or the Santa Ana.

Unassociated/anomalous: When the *insurance* supplies explained that he’d been complaining they *refused* to speak the keys.

Van Petten reported reductions in the N400 amplitude for the associated/congruent, the unassociated/congruent and the associated/anomalous conditions, relative to the unassociated/anomalous condition. She reasoned that the N400 reduction in the unassociated/congruent condition could be attributed to the sentential context and that the N400 reduction in the associated/anomalous condition was due to lexical association. Crucially, she found that the N400 amplitude reduction in the associated/congruent condition was larger than the reduction in the remaining two conditions, which she argued reflected

the additive effects of sentential context and lexical association.

Coulson et al. (2005) further explored the electrophysiological signature of associative priming in processing associated word pairs in isolation and in simple sentence contexts. Employing a lateralized half field manipulation, they additionally investigated the engagement of the two brain hemispheres subserving associative priming and the integration of sentential constraints. Their participants read pairs of associated and unassociated words in isolation. The authors reported that the second word elicited more positive ERPs in associated than in unassociated pairs. In their second experiment, the same word pairs were embedded in simple sentences. Similar to Van Petten (1993), Coulson et al. crossed the experimental factors associative relationship and sentential fit. Their results showed that processing was primarily influenced by the words' fit with the sentential contexts. However, the authors also reported subtle effects of associative priming as indexed by a more positive N400 amplitude for associated compared to unassociated pairs and a late positive component elicited by the associated but not by unassociated pairs. The influence of associative priming was particularly pronounced when the word pairs occurred in incongruent sentence contexts. With regard to hemispheric differences, Coulson and colleagues observed that after presentation to the right visual field, lexical associations had an effect in incongruous but not in congruous sentence completions. After left visual field presentation, lexical associations showed an effect in both congruous and incongruous sentences (see Beeman, 1993; Chiarello et al., 2001, for further discussion of hemispheric contributions to sentence processing).

Using visual world eye-tracking (Huettig et al., 2011, for review), Kukona et al. (2011) provided experimental evidence supporting the notion that associations and combinatorial event information jointly contribute to anticipatory processing during comprehension. The authors contrasted the influence of associative priming with the influence of event-based context on sentence comprehension. Their participants listened to sentences containing a critical verb (e.g., "arrest") such as "Toby arrests the crook" while looking at visual scenes including verb-related agents and patients (e.g., a policeman and a crook). Kukona and colleagues observed anticipatory eye movements to both agents and patients although the agent role had already been filled (Toby). They concluded that anticipatory eye gaze was influenced by simple associative relationships between the words (e.g., arrest-policeman) and by event knowledge. Crucially, associative priming showed an effect even though it conflicted with the event built up by the sentential context.

Taken together, the studies by Van Petten, Coulson et al. and Kukona et al. showed that, at the sentence level, contextual constraints exert a substantial influence on language comprehension and language-mediated anticipatory eye movements. However, there is also some evidence for the modulating influence of simple word associations (but see Paczynski and Kuperberg, 2012). The data by Coulson et al. suggest that these appear to have an impact especially when the target words are incongruent with the sentence context they appear in.

Previous investigations have shown that discourse context beyond the level of single sentences impact the comprehension process as measured using the N400 component. Hagoort et al., (2004) asked Dutch participants to read sentences such as 'Dutch trains are yellow/white/sour and very crowded'. The participants in that study knew that Dutch trains are typically yellow. Thus, the color word 'white' was a violation of their world knowledge and yielded an N400 similar to that elicited by the word 'sour'. The authors argued that comprehenders immediately integrated word meanings and world knowledge. The same lab showed in a follow-up study that the critical N400 amplitude was attenuated when the target sentence was preceded by a mitigating context. Thus, reading 'The coming world championships are one big national spectacle. The Dutch railways have painted the Dutch flag on their trains' prior to the target sentence reduced participants' N400 amplitude in response to the word 'white' in the sentence, most likely because the Dutch flag includes the color white (Hald et al., 2007; see also Van Berkum et al., 1999). Note that in these and many similar

studies the influence of the discourse context (e.g. simulation of an event) and the influence of associations between words in the discourse and the target words could not be separated. Previous attempts to do so yielded mixed results. There is consensus among researchers that simple associative relationships between words cannot fully account for the N400 amplitude modulations observed when critical target words are preceded by related discourse. However, whether or not associations contribute to discourse comprehension and if so, how strongly, is a matter of some debate.

To tease apart the effects of lexical association and the effects of discourse context, Camblin et al. (2007) orthogonally manipulated lexical associations between words and the words' congruence with the event described in the discourse. In a series of experiments recording ERPs and eye movements during reading, they investigated how lexical-level effects interacted with the effects of discourse context. The authors embedded associated and unassociated word pairs in sentences that were coherent and locally congruent. The critical words were either congruous or incongruous with the discourse-level context (see (2) for an example).

(2) **Associated/congruent:** Lynn had gotten a sunburn at the beach. Nothing she tried would help her dry and irritated skin. Lynn couldn't stop scratching her *arms* and *legs*.

Unassociated/congruent: Lynn had gotten a sunburn at the beach. Nothing she tried would help her dry and irritated skin. Lynn couldn't stop scratching her *arms* and *nose*.

Associated/incongruent: Lynn's wool sweater was uncomfortable and itchy. She fidgeted as the rough material irritated her skin. Lynn couldn't stop scratching her *arms* and *legs*.

Unassociated/incongruent: Lynn's wool sweater was uncomfortable and itchy. She fidgeted as the rough material irritated her skin. Lynn couldn't stop scratching her *arms* and *nose*.

Their analyses revealed independent effects of discourse congruence and lexical associations. Violations of discourse congruence had early and lingering effects on both ERP and eye-tracking measures, whereas the effects of associations were more fragile and particularly evident in scenarios in which the discourse context was not cohesive. Camblin et al. (2007; see Ledoux et al., 2006; Brothers et al., 2015; Otten and van Berkum, 2007, 2008, for similar conclusions) concluded that when a cohesive, congruent discourse model can be constructed, it may override associative facilitation, and effects of association "[do] not contribute to processing of words in sentences that are part of a larger discourse" (p. 126).

In a follow-up study, Boudewyn et al. (2011) used the same discourses as Camblin et al. in spoken form. In contrast to Camblin et al. (2007), they observed an interaction between lexical association and discourse congruence. Boudewyn and colleagues argued that local lexical associations and overall discourse congruence may each exert their own influence on incoming words during discourse comprehension and that these effects may be additive when the two sources of information are consistent with one another. In another study, Boudewyn and colleagues (Boudewyn et al., 2013) further demonstrated that the size of association effects during spoken discourse comprehension is subject to individual differences. Specifically, they showed that lexical associations had larger effects in individuals with lower working memory capacities as – arguably – these participants were less able to maintain discourse relationships during its unfolding. Processing in these participants was thus predominantly driven by relationships between individual words in the discourse (see also Sanford et al., 2011; Sanford and Garrod, 1998). Note, however, that the authors also observed a main effect of associations, irrespective of the individual variation between participants.

In sum, previous studies have clearly established that utilizing generalized knowledge during sentence and discourse comprehension, i. e. simulating the described event, exert powerful influences on language

comprehension (McRae and Matsuki, 2009, for review). However, the experimental results with regard to the contribution of word associations to discourse comprehension are mixed: While some studies reported significant contributions in the spoken modality (Boudewyn et al., 2011, 2013), other studies in the visual domain ascribe only a subtle influence to lexical associations, which might be overridden by the effects of discourse context (Brothers et al., 2015; Camblin et al., 2007; Otten and van Berkum, 2007, 2008). Interestingly, one condition under which associative priming has repeatedly been shown to impact language comprehension is when the critical target words are incongruent with the local contexts they are embedded in (e.g., Camblin et al., 2007; Van Petten, 1993).

In light of this finding, the possibility arises that associative priming might have contributed to the effects observed by Metusalem et al. (2012). To recap, the participants in that study read short texts about common events (e.g., a baptism) containing three kinds of target words (e.g., expected, unexpected event-related, unexpected event-unrelated). The authors found that the target words in the event-related condition that were incongruent with the local sentence context yielded an attenuated N400 amplitude relative to the event-unrelated unexpected condition. It is conceivable that the LSA post-hoc analysis carried out by Metusalem and colleagues to estimate the influence of associations was not sensitive enough to capture the interplay between sentence (in) congruence and lexical-level association effects. In order to clarify the contribution of simple associations to discourse reading and in order to gain a better understanding of the underlying mechanisms, a more direct test is needed.

1.3. The present study

In the present study, we tested the potential contributions of simple word associations to discourse reading. To that end, we first replicated Metusalem et al. (2012) context manipulation (Experiment 1). Dutch participants' EEG was recorded as they read short passages consisting of three sentences. While the first two sentences established an event context, the third sentence contained one of three target words: A highly expected word, or a word that was unexpected in the context of the third sentence but related to the overall event context, or a word that was unexpected in the context of the third sentence and unrelated to the overall event. Analyses of participants' ERPs in response to the three kinds of target words closely replicated the three-way split pattern in N400 amplitude observed by Metusalem and colleagues. This demonstrated the robustness of the results in a different language than the original English study. In Experiment 2, we asked participants to read the same target sentences as before. But now, the target sentences were preceded by two sentences, which – unlike in Experiment 1 – did not build up a coherent discourse context. However, each of the two sentences contained a word (prime, hereafter) that was part of the event-establishing sentences in Experiment 1 and that was associatively related to the unexpected event-related target word in that discourse. Using this manipulation, we minimized the event knowledge that participants could extract from the two introductory sentences while keeping part of the associated lexical input the same as in Experiment 1. Thereby, we substantially reduced the possibility of participants using event knowledge to generate predictions about the upcoming target. As in Experiment 1, the primes appeared in grammatically well-formed sentences.

If the N400 amplitude reduction in the unexpected event-related condition in Experiment 1 (and in Metusalem et al., 2012, Experiment 1) was partly due to associative priming, we should observe a similar pattern of results in Experiment 2. Specifically, we predicted a significant difference between the N400 amplitude elicited by the unexpected event-related condition and the N400 amplitude elicited by the unexpected event-unrelated condition. Moreover, we predicted both unexpected conditions to differ significantly from the expected condition. Such a pattern would be consistent with the notion that simple

associative relationships between words modulated the critical N400 component in the previous experiments. It would also be consistent with the notion that associations contribute to discourse comprehension. If on the other hand, the effect in the previous experiments was primarily driven by the activation of event knowledge, the event-related and event-unrelated unexpected conditions should now elicit N400 components of similar amplitude (as participants can rely less on event simulation to activate the target), with both being more negative than the N400 elicited in the expected condition.

Note that Metusalem and colleagues linked the reduced N400 amplitude in the unexpected, event-related condition to the notion that readers utilize event knowledge to predict referents that are likely to be mentioned in the described event. Similarly, many of the ERP studies reviewed above have discussed their findings in the context of predictive language comprehension. Indeed, we believe it is likely that both event simulation and associations are mechanisms contributing to predictive language processing. However, given recent methodological debates, we cannot label the behavior reported by Metusalem et al. conclusively as predictive, as the N400 component in that study was measured on the target words and not before (see Baggio and Hagoort, 2011; Huettig and Mani, 2016; Kutas and Federmeier, 2011; Mantegna et al., 2019; Nieuwland et al., 2018, for discussion). We return to this issue in the General Discussion.

2. Experiment 1

2.1. Method

2.1.1. Participants

Thirty-one members of the subject panel of the MPI (five male, mean age = 21, $SD = 2$) took part in Experiment 1. All were native speakers of Dutch, right-handed, and did not report any history of learning or reading disabilities or neurological or psychiatric disorders. The participants were paid for participation. The ethics board of the Faculty of Social Sciences of the Radboud University in Nijmegen approved the study. One participant had to be excluded from the analysis due to an experimental error.

2.1.2. Stimuli

We started by translating the English materials used by Metusalem et al. (2012) into Dutch. Each of the 72 experimental items consisted of a short discourse that consisted of three sentences (see Table 1 for sample stimuli used in Experiment 1). The first two sentences established the

Table 1
Sample stimuli used in Experiment 1 and 2.

| Experiment | Lead-in sentences | Target sentence |
|------------|--|---|
| 1 | De ouders waren erg blij met hun pasgeboren dochter. Een van de eerste dingen die ze hebben gedaan is haar laten dopen in hun kerk. | De baby hield ervan om in bad te gaan dus ze lachte toen ze werd besprenkeld met water/pastoor/tandarts op haar voorhoofd. |
| | <i>The parents were very excited about their new baby girl. One of the first things they did was to get her baptized in their church.</i> | <i>The baby liked baths, so she smiled when she was sprinkled with water/priest/dentist on her forehead.</i> |
| 2 | Dat Peter en Claudia de jongen lieten dopen, was het onderwerp van het gesprek. Telkens weer bleek het plaatje van de kerk in het geschiedenisboek zijn aandacht te trekken. | De baby hield ervan om in bad te gaan dus ze lachte toen ze werd besprenkeld met water/pastoor/tandarts op haar voorhoofd. |
| | <i>That Peter and Claudia had the boy baptized, was the subject of the conversation. Time and again the image of the church in the history book appeared to get his attention.</i> | <i>The baby liked baths, so she smiled when she was sprinkled with water/priest/dentist on her forehead.</i> |

event context. The third sentence contained the target words. The comprehension questions used by Metusalem and colleagues were also translated. Five native speakers of Dutch checked the qualities of the translations. The subsequent rating studies were the same as described by Metusalem and colleagues and were carried out to ensure that the Dutch materials were comparable to those of the original study.

Cloze probability. We pre-tested the predictability of the expected target words within each discourse. Thirty-one native speakers of Dutch (five male; mean age = 22, $SD = 3$), none of whom participated in the main experiments or in any of the other rating studies, carried out a cloze probability rating study. The study was conducted online using a tool for web experiments developed by the technical group of the Max Planck Institute for Psycholinguistics. The 72 translated discourses were presented to the participants in random order, one at a time, up to the word preceding the expected target. Participants were instructed to read each discourse carefully and provide the word they thought would be the best continuation of the sentence. Cloze probability was the proportion of participants who provided a particular response for a given discourse (Taylor, 1953). For each discourse, we selected the word with the highest cloze probability as the expected target of that discourse. On eleven items another word than the direct Dutch translation of the English expected target was deemed the most likely continuation of the third sentence. Due to very low cloze probability, we had to exclude six items, which were part of the original material set. The mean cloze probability of the expected targets in the remaining 66 items was 0.67 (range = 0.19 - 1; $SD = 0.23$; mean cloze probability in Metusalem et al., 2012, was 0.81, based on 30 participants and 72 items).

Event association. In another web-based rating study, we instructed 49 native speakers of Dutch (ten male; mean age = 20, $SD = 2$), who did not take part in the main experiments or other rating studies, to read the discourses containing the expected targets and try to think of persons and objects that were likely to be part of the described scenarios but were not mentioned. The discourses were randomized for each participant and presented one at a time. Participants were asked to write down minimally three and up to five event-associated persons and objects. Participants provided 13,180 out of 16,170 possible associations (49 participants x 66 items x 5 associations). Five-hundred sixteen of these (4%) had to be removed because they were not nouns. Event-related targets were selected according to a weighting system that was based on the order of mention of participants' responses (five points for the first association, four points for the second, etc.). The highest scoring association that was not listed in the cloze probability rating was selected as event-related target.

The Dutch language - unlike English - differentiates between common and neuter grammatical gender of nouns. Gender is reflected in determiners (common gender: *de* vs. neuter gender: *het*; both are equivalent to the English *the*) and in inflectional marking on pronominal adjectives (e.g., common gender: *een goede auto*, a good car vs. neuter gender: *een goedø boek*, a good book). When the highest scoring event-related target had a different gender from the expected target noun, and when gender was marked by a determiner or adjective, the highest scoring event-related noun with the same gender as the expected target was chosen. The maximum possible score for an item was 245 (49 participants x score of 5). Across the 66 discourses, the mean relatedness score for the event-associated targets was 75 (range = 16 to 222, $SD = 48$; mean event-related score in Metusalem et al., 2012, was 92.4, based on 45 participants and 72 items¹).

¹ Note that the cloze probability of the expected target words and the event-relatedness scores for the unexpected event-related targets were a bit lower than in the original study. The lower cloze probability is very likely connected to the fact that we had to exclude six items from the original set. The event-relatedness scores were lower than in Metusalem et al.'s study because sometimes low event associations had to be selected as target words to match the expected target words' gender marking.

The event-unrelated targets were generated by shuffling the event-related targets across discourses such that event-related and event-unrelated targets consisted of the same lexical items. Therefore, lexical factors such as length and word frequency were the same across the two conditions. We split the 66 experimental items into three lists each containing 22 discourses. Each discourse appeared once in each list and once in each condition across the three lists. We minimized the variability across the lists by matching the three lists on the following variables: mean cloze probability, log frequency, and orthographic length of the expected targets; mean event-relatedness score, log frequency, and orthographic length of the event-related targets. Finally, we shuffled the event-related targets across the discourses within each rotation group to obtain the event-unrelated targets. The shuffling was done such that the event-related and event-unrelated targets within each discourse were matched for animacy and concreteness. We checked that in case of overt gender marking the event-unrelated targets had the same gender as the expected targets and the event-related targets. The results of all rating studies are summarized in Table 2.

The association strength between the expected and the unexpected targets was checked to assess the possibility that potential reductions in the amplitude of the N400 were driven by strong associative connections between the target words (cf. Federmeier and Kutas, 1999; Kleiman, 1980). We used the Dutch free association database by De Deyne et al. (2013) to determine how strongly the expected targets (cues) and the event-related and event-unrelated targets (responses) were related. Fifty-nine of the 66 expected targets were listed in the database. The mean associative strength for these items was 0.0057 ($SD = 0.014$) for the event-related targets and was 0.0011 ($SD = 0.007$) for the event-unrelated targets, which amounts to one response per 175 participants and one response per 909 participants, respectively. De Deyne et al.'s database lists responses of the first 100 participants who read a particular cue word. Thus, the association strength between the expected targets and both kinds of unexpected targets was rather low.

Twenty-two filler items of the same structure as the experimental items were constructed to ensure an equal number of trials containing anomalous and non-anomalous targets. Each discourse was followed by a comprehension question about the just-read scenario. Yes and No responses occurred equally often.

2.1.3. Procedure

Participants were randomly assigned one of the three lists and were tested individually in a dimly lit room. They were seated in a relaxed position in front of a 19-inch CRT screen. We told them that they were going to participate in a reading comprehension experiment, which consisted of a number of short discourses. Participants were instructed to read each of these discourses carefully in order to be able to answer

Table 2

Rating results for the three rotation groups and the stimuli set overall. Word frequencies were taken from the SUBTLEX-NL corpus of Dutch subtitles (Keuleers et al., 2010).

| | List 1 | List 2 | List 3 | Overall |
|--------------------------------|--------|--------|--------|---------|
| <i>Expected targets</i> | | | | |
| Cloze probability | .66 | .68 | .69 | .68 |
| Word frequency | 71 | 64 | 70.9 | 68.63 |
| Orthographic length | 6.6 | 6.1 | 6 | 6.23 |
| <i>Event-related targets</i> | | | | |
| Cloze probability | .00 | .00 | .00 | .00 |
| Word frequency | 92.9 | 91 | 106 | 96.63 |
| Orthographic length | 6.4 | 5.9 | 6.4 | 6.23 |
| Event relatedness score | 75.5 | 79.9 | 68.1 | 74.5 |
| <i>Event-unrelated targets</i> | | | | |
| Cloze probability | .00 | .00 | .00 | .00 |
| Event relatedness score | 0 | 0 | 0 | 0 |

the questions following the discourses. They were instructed to move or blink as little as possible while reading the third sentence. The order of experimental and filler trials was randomized in the beginning of the experiment.

The trial parameters were identical to those used by Metusalem et al.: Participants were presented with the first two sentences of a discourse in the middle of the screen. Once they had read and understood the two sentences, they pushed a button on the response device to advance to the third sentence. A red fixation cross in the middle of the screen appeared for 1000 ms and cued the beginning of the third sentence which was presented word by word (i.e., rapid serial visual presentation, RSVP). The stimulus onset asynchrony for the words was 350 ms, divided into 200 ms presentation of the word and 150 ms inter-stimulus interval. Directly following the last word in the sentence, the comprehension question appeared in the middle of the screen requiring participants to provide a Yes or No response. Responses were given by pushing a button on the response device using either the right or left thumb. The left-right-Yes-No button assignment was counter-balanced across participants. The experiment consisted of five blocks, which were separated by pauses. The first block consisted of 20 trials; the remaining four blocks consisted of 17 trials. The entire session took a little less than 2 h.

2.1.4. EEG recording and analysis

EEG was recorded continuously from 26 active Ag/AgCl electrodes mounted in a cap according to the 10–20 system (Klem et al., 1999). The signal was amplified by a Biosemi active amplifier with a bandpass filter of 0.016–100 Hz, sampled at a frequency of 250 Hz and referenced online to the left mastoid. Four additional electrodes were used to monitor participants' horizontal and vertical eye movements and blinks (see Table 3, for an overview of the electrode groupings for distributional analyses). All electrode impedances were kept below 5 K Ω .

The ERP analysis was carried out on the experimental items of each participant individually, using Brain Vision Analyzer (version 2.0). Participants' EEG data were re-referenced offline to the average of the left and right mastoids and filtered again (highpass = 30 Hz, lowpass = .01 Hz). The EEG was time-locked to the onset of the three kinds of targets in the recording. Semi-automatic artifact rejection was used to exclude trials containing drifts, blinks, and muscle tension. The overall percentage of trials excluded due to artifact contamination was 8%, which was evenly distributed over the three conditions (expected targets = 7%, event-related targets = 9%, event-unrelated targets = 8%). Ocular correction was applied based on the average of each participant's vertical and the average of their horizontal eye electrodes (cf. Gratton et al., 1983). As in the original study by Metusalem et al., by-participant averages for the three conditions were calculated, relative to a 500-ms pre-stimulus baseline window. The individual participant averages were then averaged together yielding a grand average ERP for each condition.

The analysis of participants' accuracy on the comprehension questions suggested that they read and understood the discourses (mean accuracy = 97%, $SD = 3\%$).

Table 3

Summary of electrode groupings for the distributional analyses in Experiments 1 and 2.

| | | |
|-------------|------------|---------------------------------------|
| Hemisphere | Right | F4, FC2, FC6, C4, CP2, CP6, P4, O2 |
| | Left | F3, FC1, FC5, C3, CP1, CP5, P3, O1 |
| Laterality | Lateral | F3, FC5, CP5, P3, F4, FC6, CP6, P4 |
| | Medial | FC1, C3, CP1, O1 FC2, C4, CP2, O2 |
| Anteriority | Prefrontal | F3, F4, FC1, FC2 |
| | Frontal | FC5, FC6, C3, C4 |
| | Parietal | CP1, CP2, CP5, CP6, |
| | Occipital | P3, P4, O1, O2 |

3. Results

Fig. 1 displays the grand average ERPs elicited by the target words in Experiment 1 for the 26 scalp electrodes. We plot 1000 ms post-stimulus onset and 500 ms pre-stimulus baseline. Visual inspection suggests no differences between the three waveforms prior to the onset of the target word presentation. With regard to the amplitude of the N400, a three-way split very similar to the pattern observed by Metusalem et al. arose around 300 ms after stimulus onset, extending to roughly 600 ms after stimulus onset. The N400 amplitude elicited by the expected targets is positive at the majority of the electrodes. The N400 amplitude elicited by the unexpected event-unrelated targets is the most negative going deflection of the three conditions. Fig. 1 indicates that the amplitude of the N400 elicited by the unexpected event-related targets lay between these two conditions (see Fig. 2, for the three-way split in N400 amplitude on a central-parietal electrode, Pz).

Following Metusalem et al.'s approach, we analyzed the N400 amplitude differences across the three conditions using a repeated measures ANOVA with three levels of Condition and 26 levels of Electrode. Mean ERP amplitudes were calculated for the time window from 300 to 500 ms after target word onset. The analysis yielded main effects of Condition, $^2 F(2,58) = 30.421$, $\mathcal{E}_{GG} = 0.924$, $p < .001$, a main effect of Electrode, $F(25,725) = 9.637$, $\mathcal{E}_{GG} = 0.140$, $p < .001$, and a Condition-by-Electrode interaction, $F(50,1450) = 10.325$, $\mathcal{E}_{GG} = 0.18$, $p < .001$. The subsequent planned comparison confirmed the visual inspection and thus the replication of Metusalem et al.'s findings. The N400 amplitude elicited by the event-related targets was between the amplitudes elicited by the expected and the unexpected event-unrelated targets. That is, N400 amplitudes in the event-unrelated condition were significantly greater (i.e., more negative-going) than those in the event-related condition ($F(1,29) = 17.461$, $\mathcal{E}_{GG} = 1$, $p < .001$; interaction with electrode: $F(25,725) = 5.225$, $\mathcal{E}_{GG} = 0.28$, $p < .001$) and significantly smaller than those in the expected condition ($F(1,29) = 15.981$, $\mathcal{E}_{GG} = 1$, $p < .001$; interaction with electrode: $F(25,725) = 10.573$, $\mathcal{E}_{GG} = 0.225$, $p < .001$).

Fig. 1 further suggests that the three-way N400 split pattern was widespread across the scalp, but was most prominently expressed over medial parietal-occipital sites. Fig. 3 shows difference waves for the unexpected conditions (event-related minus expected and event-unrelated minus expected) demonstrating the size of the N400 effect at all 26 electrodes (see Fig. 4, for scalp topographies of the mean amplitudes in the 300–500 ms time window for these conditions). To assess the exact topographic distribution, mean ERP amplitude difference scores between each unexpected condition and the expected condition during the 300–500 ms time window were submitted to a repeated measures ANOVA with two levels of Difference (event-related minus expected vs. event-unrelated minus expected), two levels of Hemisphere (left vs. right), two levels of Laterality (lateral vs. medial) and four levels of Anteriority (pre-frontal vs. frontal vs. parietal vs. occipital). The model revealed main effects of Difference ($F(1,29) = 20.681$, $\mathcal{E}_{GG} = 1$, $p < .001$), Laterality ($F(1,29) = 16.914$, $\mathcal{E}_{GG} = 1$, $p < .001$) and Anteriority ($F(3,87) = 4.18$, $\mathcal{E}_{GG} = 0.464$, $p < .035$), as well as interactions between Difference and Laterality ($F(1,29) = 5.093$, $\mathcal{E}_{GG} = 1$, $p < .032$) and Difference, Laterality and Anteriority ($F(3,87) = 15.454$, $\mathcal{E}_{GG} = 0.841$, $p < .001$). The effect of Hemisphere was not statistically reliable ($F(1,29) = 2.486$, $\mathcal{E}_{GG} = 1$, $p = .126$). We explored the significant effects by means of Bonferroni-corrected post-hoc tests. The main effect of Difference shows once more that the N400 components elicited by the unexpected event-unrelated target words were more negative-going than the one elicited by the unexpected event-related target words (p

² We report p-values for Greenhouse-Geisser epsilon-adjusted degrees of freedom (Greenhouse and Geisser, 1959), the unadjusted degrees of freedom, and the value of the Greenhouse-Geisser epsilon for F-tests with more than one degree of freedom in the numerator.

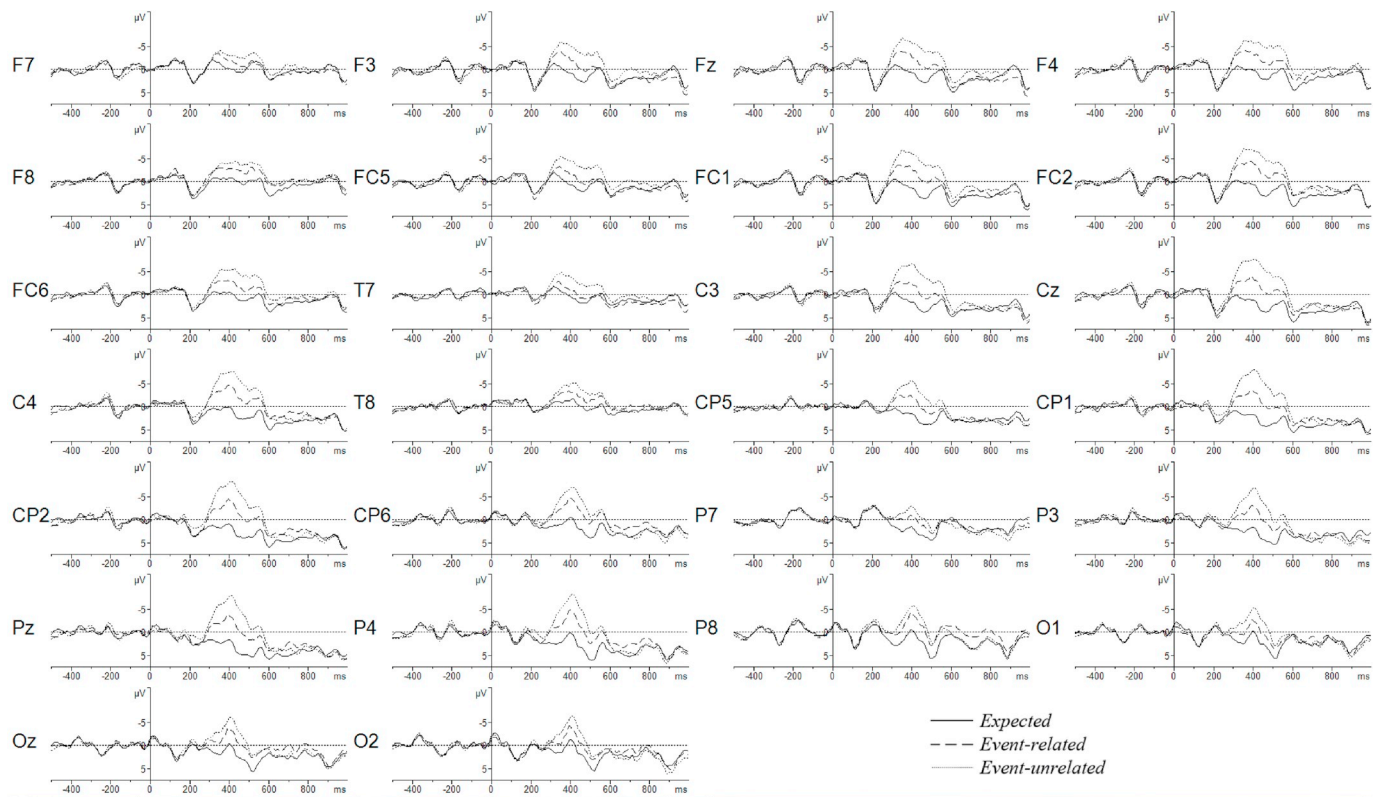


Fig. 1. Grand average ERPs elicited by the target words in the three conditions in Experiment 1. Time zero refers to the onset of the target word presentation. Negative voltage is plotted up.

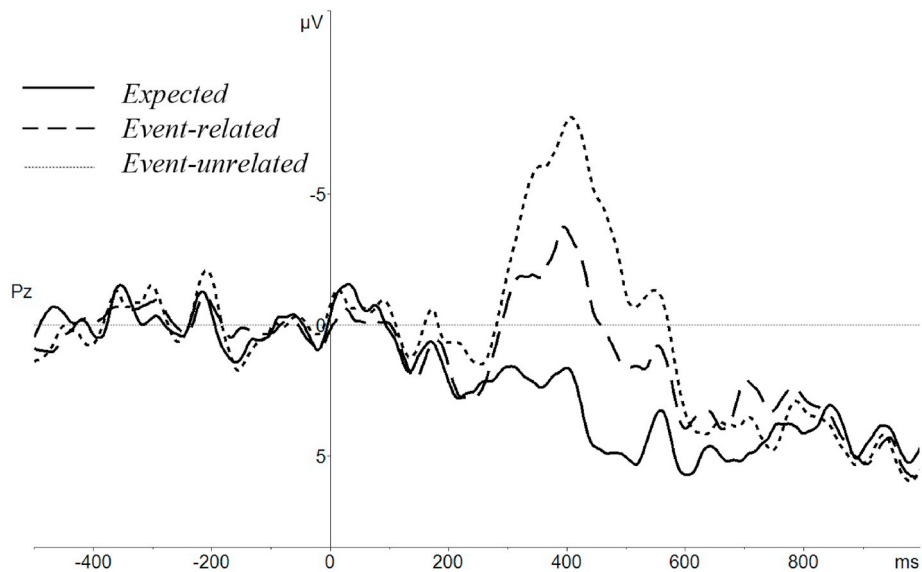


Fig. 2. Grand average ERPs at the midline parietal electrode (Pz) in Experiment 1.

< .001). With regard to the factor Laterality, N400 amplitudes appear to be generally more negative over medial than over lateral sites ($p < .001$). The main effect of Anteriority indicates that N400 amplitudes were more negative over parietal-occipital electrodes (frontal vs. parietal: $p < .001$; parietal vs. occipital: $p = .003$). The Difference-by-Laterality interaction revealed that the difference in N400 amplitude between event-related and event-unrelated conditions was more positive over medial than over lateral sites ($p < .001$).

Fig. 1 and 2 suggest that the deflections elicited by the three kinds of

target words may additionally differ in time windows preceding and following that of the N400 component. More specifically, Fig. 1 indicates that, around 200 ms post-stimulus onset, the deflections elicited in the expected condition and in the unexpected event-related condition were more positive than the deflection elicited in the unexpected event-unrelated condition. Previous research has linked variation in ERP amplitude during that time window (the so-called ‘P200 component’) to variance in the level of expectancy for particular lexical items. Federmeier and Kutas (2002) proposed that during reading language users

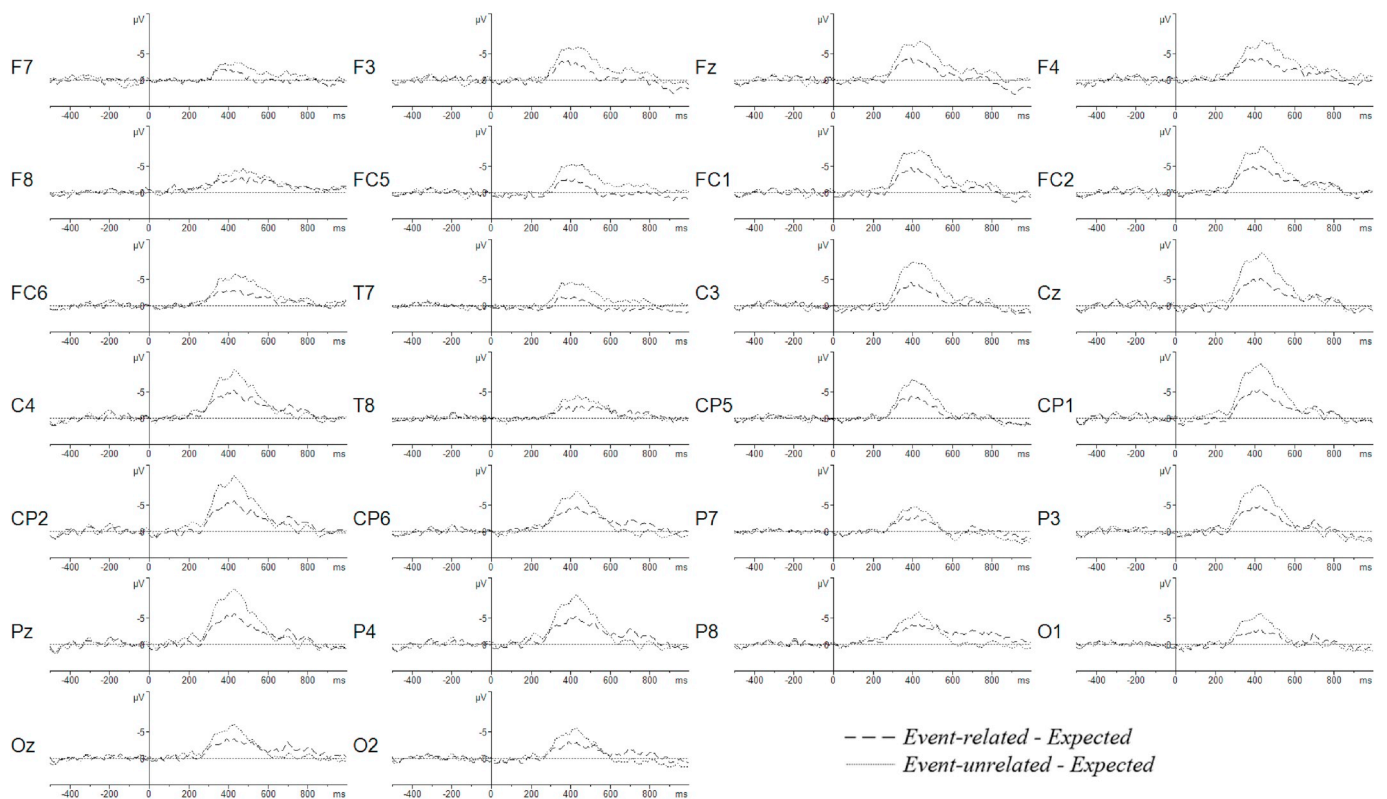


Fig. 3. Difference waves reflecting the size of N400 effects in the event-related and event-unrelated conditions in Experiment 1.

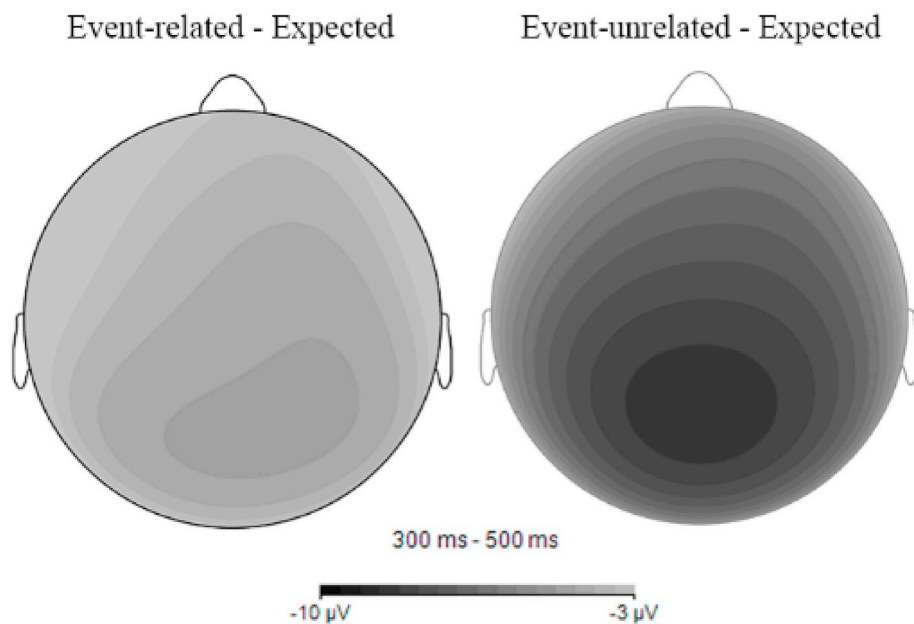


Fig. 4. Scalp topographies of the N400 effects in the event-related and event-unrelated conditions from Experiment 1. The left plot reflects the N400 effect for the event-related targets, and the right the event-unrelated targets. Values correspond to mean amplitude, 300–500 ms post-stimulus onset in the respective condition.

rely on contextual information in a sentence to prepare for the visual analysis of an upcoming stimulus. Roughly 200 ms post-stimulus onset, expected words are assumed to elicit more positive-going amplitudes than unexpected words. Furthermore, similar to Metusalem et al. (2012, Experiment 1), Fig. 1 suggests that the deflection elicited by the unexpected event-unrelated condition was more negative over frontal regions (with a slight preference for the left hemisphere) than the deflections elicited by expected and unexpected event-related

conditions during the time window following the N400 component. Moreover, the deflection elicited by the expected target appears to be more positive than the deflections elicited by the other two conditions over right-hemispheric fronto-parietal regions.

Although the main focus of the present study was on examining differences in N400 amplitude across the three conditions, we conducted two additional analyses. We entered mean amplitudes from 150 to 250 ms and from 500 to 900 ms, respectively, to repeated measures ANOVAs

with three levels of Condition and 26 levels of Electrode each. The results of the P200 analysis showed a main effect of Electrode ($F(25,725) = 24.087$, $\mathcal{E}_{GG} = 0.176$, $p < .001$) but neither a main effect of Condition ($F(2,58) = 1.066$, $\mathcal{E}_{GG} = 1$, $p > .3$) nor an interaction between Condition and Electrode ($F(50,1450) = 1.47$, $\mathcal{E}_{GG} = 0.264$, $p = .16$). The results of the second analysis, concerning the 500–900 ms time window, likewise showed a main effect of Electrode ($F(25,725) = 17.29$, $\mathcal{E}_{GG} = 0.215$, $p < .001$) but not Condition ($F(2,58) = 1.859$, $\mathcal{E}_{GG} = 0.897$, $p = .17$). The Electrode-by-Condition interaction, however, was statistically reliable ($F(50,1450) = 3.163$, $\mathcal{E}_{GG} = 0.209$, $p = .001$). The presence of these effects is interesting, in particular because they closely resemble the effects reported by Metusalem and colleagues.

4. Discussion

The results show the three-way split pattern in N400 amplitude as observed by Metusalem et al. (2012). That is, the N400 amplitude elicited by the expected targets was more positive than the N400 amplitude elicited by the unexpected event-related targets and was more positive than the N400 elicited by the unexpected event-unrelated targets. Crucially, the N400 amplitude elicited by the unexpected event-related targets was less negative than the N400 elicited by the unexpected event-unrelated targets. These findings demonstrate the robustness of the effect in a different language than the original English study. Our analyses revealed that the N400 attenuation in the unexpected event-related condition was most strongly visible over medial parietal-occipital electrodes. In sum, the results are consistent with the notion that event simulation contributes to discourse reading.

4.1. Experiment 2

Given the close replication of the original experiment, we could assess the contribution of associative priming to the attenuated N400 amplitude in the unexpected event-related condition in Experiment 1. In Experiment 2, participants read short paragraphs, each consisting of three sentences. In each paragraph, the same passage-final sentence as in Experiment 1, including the same three kinds of target words, was used (e.g., ‘The baby liked baths, so she smiled when she was sprinkled with *water/priest/dentist* on her forehead’). We selected two prime words (e.g., ‘baptized’, ‘church’) from the event-establishing sentences in Experiment 1 that, from all the words in these sentences, were most strongly associated with the unexpected event-related target words (e.g., ‘priest’). The two primes were each placed in a neutral carrier sentence (see Table 1 for an example). We controlled that none of the other words in the two carrier sentences was associated with any of the target words in that paragraph or with the primes. An additional rating study ensured that reading the two introductory sentences in a given paragraph substantially reduced the possibility of simulating a cohesive event. Yet, the primes were embedded in a meaningful syntactic environment. Thereby, we also minimized potential effects of strategic processing as the associative relationship between the critical targets and the primes became less obvious.

If associative priming contributed to the N400 amplitude reduction in the unexpected event-related condition in Experiment 1, event-related targets in Experiment 2, which were preceded by associatively related primes in isolated carrier sentences, should elicit an attenuated N400 amplitude, relative to the N400 amplitude elicited by the unexpected event-unrelated targets. Note that we generally predicted smaller effects in Experiment 2. As discussed in the Introduction, previous studies showed that discourse context exerts a powerful influence on the amplitude of the N400 component, whereas the effects of associations are rather subtle. As we minimized the influence of discourse context, we expected somewhat smaller differences between the N400 amplitudes elicited by the three kinds of target words. If, alternatively, the graded N400 pattern in the previous experiment was primarily driven by event knowledge, event-related and event-unrelated unexpected target

words should elicit N400 components with similar amplitudes.

5. Method

5.1. Participants

Thirty-three volunteers (three male, mean age = 22, $SD = 3$) who did not take part in Experiment 1 or in any of the rating studies participated in Experiment 2. All were right-handed, native speakers of Dutch and did not report any history of learning or reading disabilities or neurological or psychiatric disorders. Due to an experimental error, the log-files of two participants were not saved. Another participant was excluded post-hoc due to too much noise in the EEG signal.

5.2. Stimuli and procedure

In each experimental item of Experiment 1, we replaced the two event-establishing sentences with two isolated sentences. Each of these contained a prime word. Prime words were selected from the event-establishing sentences and were most strongly associated with the unexpected event-related target word in that item. The association strength between a chosen prime word (e.g. verb, noun, adjective, preposition) and the three target words was determined using the Dutch free association database by De Deyne and colleagues (De Deyne et al., 2013). In case of inflections, the base form of that word was looked up. Thirteen (out of 132) prime words were not listed in the database. The average association strength between the remaining 119 primes and the unexpected event-related target words was 0.0282 ($SD = 0.0557$; one response in 35 participants). The average association strength between the 119 primes and the expected targets and the unexpected event-unrelated targets was 0.013 ($SD = 0.0351$; one response in 77 participants) and 0.0002 ($SD = 0.0013$; one response in 5000 participants), respectively. The primes were embedded in neutral non-predictable carrier sentences and appeared in the same inflectional form as in the event-establishing sentences. None of the other words in the two carrier sentences was associatively related to the three target words within a given paragraph (see Table 1 for a sample stimulus used in Experiment 2).

Discourse coherence rating. We assessed to which extent the two carrier sentences and the target sentence built up a coherent discourse context by asking 36 Dutch speaking participants (mean age = 23, $SD = 4$), who did not take part in one of the main experiments or the rating studies, to provide ratings in a pen-and-paper test. To that end, we distributed both versions of an item, coherent (Experiment 1) and incoherent (Experiment 2), over two lists such that one item only occurred once on one list and that both lists featured the same number of coherent and incoherent items. Participants were randomly assigned one list and saw one item at a time, presented on a PowerPoint slide. They were instructed to rate on a scale from 1 (incoherent) to 10 (perfectly coherent) how well the target sentence fitted with the first two preceding sentences. Note that we included the coherent items from Experiment 1 to have a baseline against which to compare the respective incoherent version. The average rating for the coherent discourses was 9.1 ($SD = 0.6$); the average rating for the incoherent discourses was 4.0 ($SD = 1.5$). A *t*-test confirmed that the ratings for both item versions differed significantly ($t(65) = 27.24$, $p < .001$). Taken together, the results of the rating study suggest that, as expected, discourse coherence differed strongly across experiments.

The filler items used in Experiment 1 were edited by shuffling the event-establishing sentences and the target sentences across all paragraphs. By doing so, completely unrelated sentences were paired, which did not form a coherent discourse either. Finally, we created new comprehension questions for all experimental and filler items. Within each item, the question focused on the contents of either the first or the second carrier sentence. The ratio of Yes and No responses was balanced. Apart from these changes, the procedure was the same as in Experiment

1. That is, participants were first presented with the two introductory sentences, presented centrally on the computer screen. After having read these sentences, they pressed the space bar and the target sentence was presented in RSVP. The analysis was identical to that of Experiment 1.

6. Results and discussion

The accuracy analysis of their responses to the comprehension questions showed that participants read the carrier sentences carefully and understood the content (mean accuracy = 90%, *SD* = 5%).

Fig. 5 plots the grand average ERPs elicited by the three conditions in Experiment 2 for all 26 scalp electrodes (Fig. 6, for close-up of Pz). Visual inspection suggests no differences between the three waveforms prior to the onset of target word presentation. However, at around 300 ms after target word onset, the lines diverge, extending to around 600 ms post-stimulus onset. The N400 amplitude elicited by the unexpected event-unrelated targets was less negative than in Experiment 1. Moreover, the N400 amplitude elicited by the expected targets was more negative than in Experiment 1. This was expected, and both effects are most likely connected to the lower predictability of the target words in the expected condition. As we minimized the amount of event knowledge in the paragraphs by replacing the three coherent sentences in one paragraph with three isolated sentences, the expected targets were less predictable than in Experiment 1 and thus elicited a more negative-going N400 amplitude. In turn, as the expected target words were less predictable, the degree of expectation violation in response to the unexpected event-unrelated target words was smaller, too, hence the less negative N400 amplitude. Interestingly, visual inspection suggests that the N400 amplitude elicited by the unexpected event-related target words was similar to the amplitude elicited by the same condition in Experiment 1.

To analyze the effects statistically, mean ERP amplitudes were submitted to a repeated measures ANOVA with three levels of Condition and 26 levels of Electrode. This analysis yielded main effects of

Condition, $F(2,58) = 12.252$, $\mathcal{E}_{GG} = 0.813$, $p < .001$ and Electrode, $F(2,58) = 4.772$, $\mathcal{E}_{GG} = 0.137$, $p < .001$ and a Condition-by-Electrode interaction $F(50,1450) = 1.848$, $\mathcal{E}_{GG} = 0.193$, $p = .041$. The planned comparison between event-related and event-unrelated unexpected conditions revealed the statistical reliability of the critical difference, $F(1,29) = 5.217$, $\mathcal{E}_{GG} = 1$, $p = .03$ (interaction with electrode: $F(25,725) = 1.044$, $\mathcal{E}_{GG} = 0.244$, $p > .3$; see Fig. 7 for difference waves between the unexpected and the expected conditions). The unexpected event-related condition also differed from the expected condition ($F(1,29) = 9.465$, $\mathcal{E}_{GG} = 1$, $p = .005$; interaction with electrode: $F(25,725) = 2.993$, $\mathcal{E}_{GG} = 0.229$, $p = .01$).

We analyzed the scalp topographies of the N400 effects for the event-related and event-unrelated targets according to the same procedure as in Experiment 1. This yielded a significant main effect of Difference ($F(1,29) = 5.114$, $\mathcal{E}_{GG} = 1$, $p = .030$). While the main effect of Anteriority was trending towards significance ($F(3,87) = 2.348$, $\mathcal{E}_{GG} = 0.474$, $p = .123$), none of the other main effects or interactions were statistically reliable (all $p > .4$). Bonferroni-corrected post-hoc tests showed that N400 amplitudes were most negative over medial frontal-parietal electrodes (pre-frontal vs. occipital: $p = .05$, frontal vs. occipital: $p = .024$, parietal vs. occipital: $p = .029$; all other comparisons: $p > .3$; Fig. 8).

To compare effect sizes in Experiment 1 and 2 directly, we submitted difference scores between event-related and expected conditions and event-unrelated and expected conditions, respectively, of both experiments, to a repeated measures ANOVA with the factors Difference (event-related minus expected conditions vs. event-unrelated minus expected conditions; within-participants), Electrode (26 levels; within-participants) and Experiment (Experiment 1 vs. Experiment 2; between-participants). We obtained significant main effects of Difference ($F(1,29) = 22.671$, $\mathcal{E}_{GG} = 1$, $p < .001$), Electrode ($F(25,725) = 15.038$, $\mathcal{E}_{GG} = 245$, $p < .001$) and Experiment ($F(1,58) = 10.03$, $p = .002$), and interactions between Electrode and Difference ($F(25) = 5.334$, $\mathcal{E}_{GG} = 0.297$, $p < .001$), Electrode and Experiment ($F(25,1450) = 5.1$, $p < .001$) as well as Difference and Experiment ($F(1,1) = 6.981$, $p <$

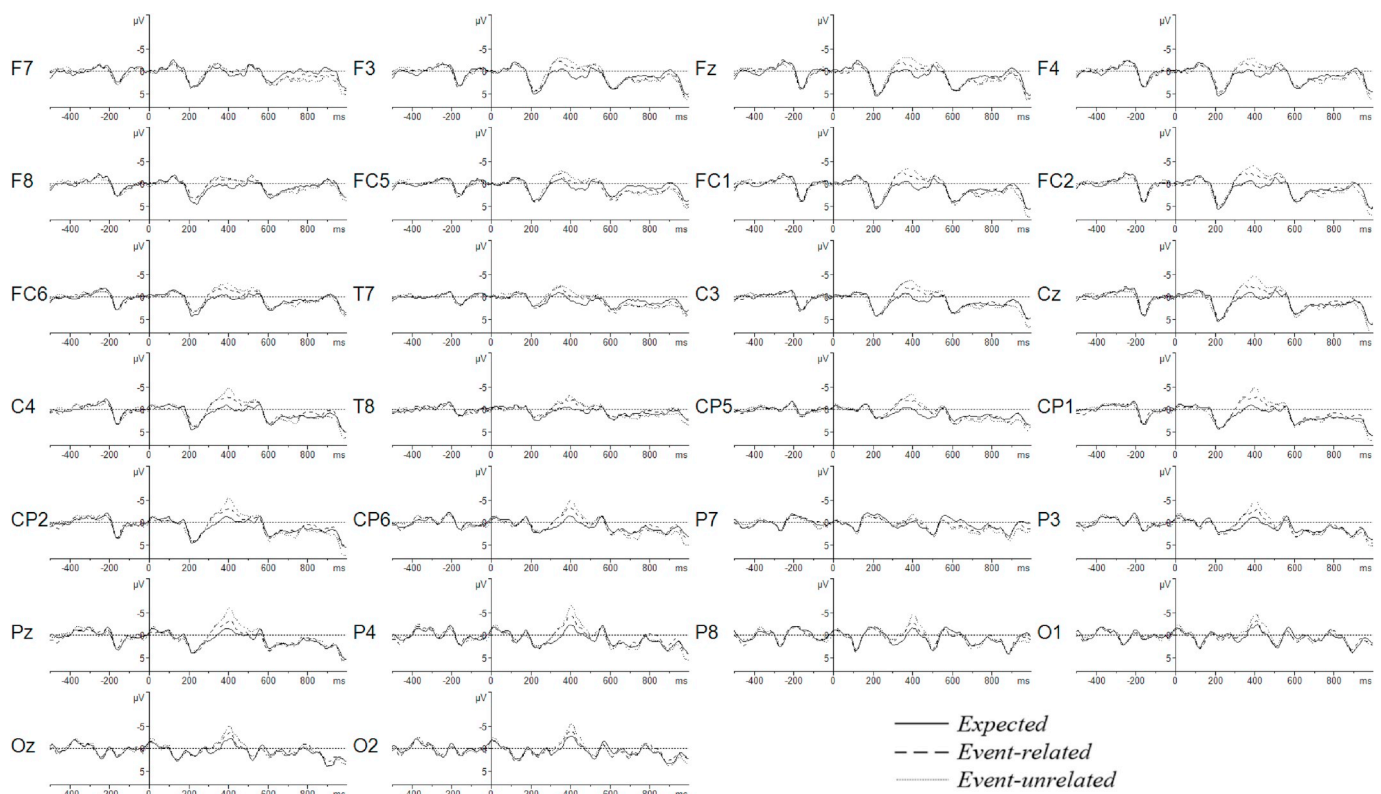


Fig. 5. Grand average ERPs elicited by the target words in the three conditions in Experiment 2.

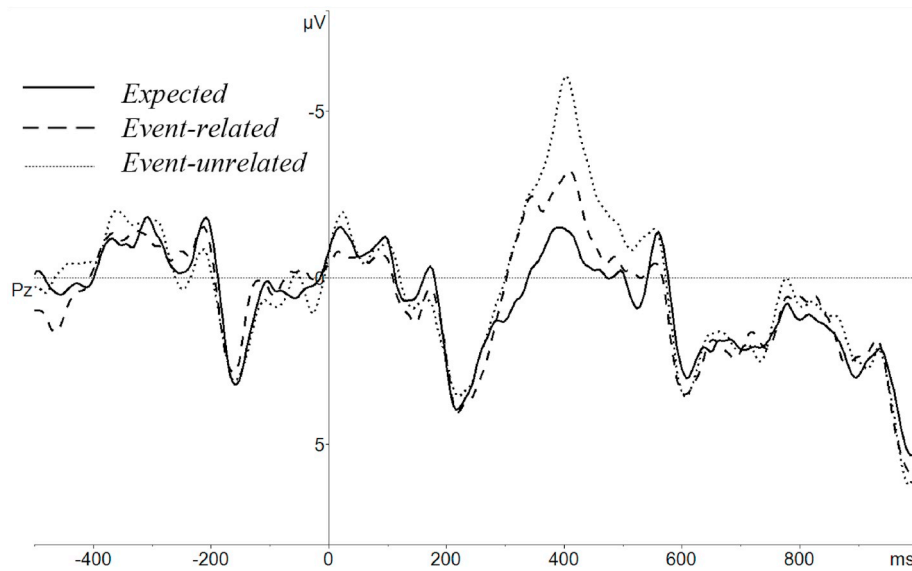


Fig. 6. Grand average ERPs at the midline parietal electrode (Pz) in Experiment 2.

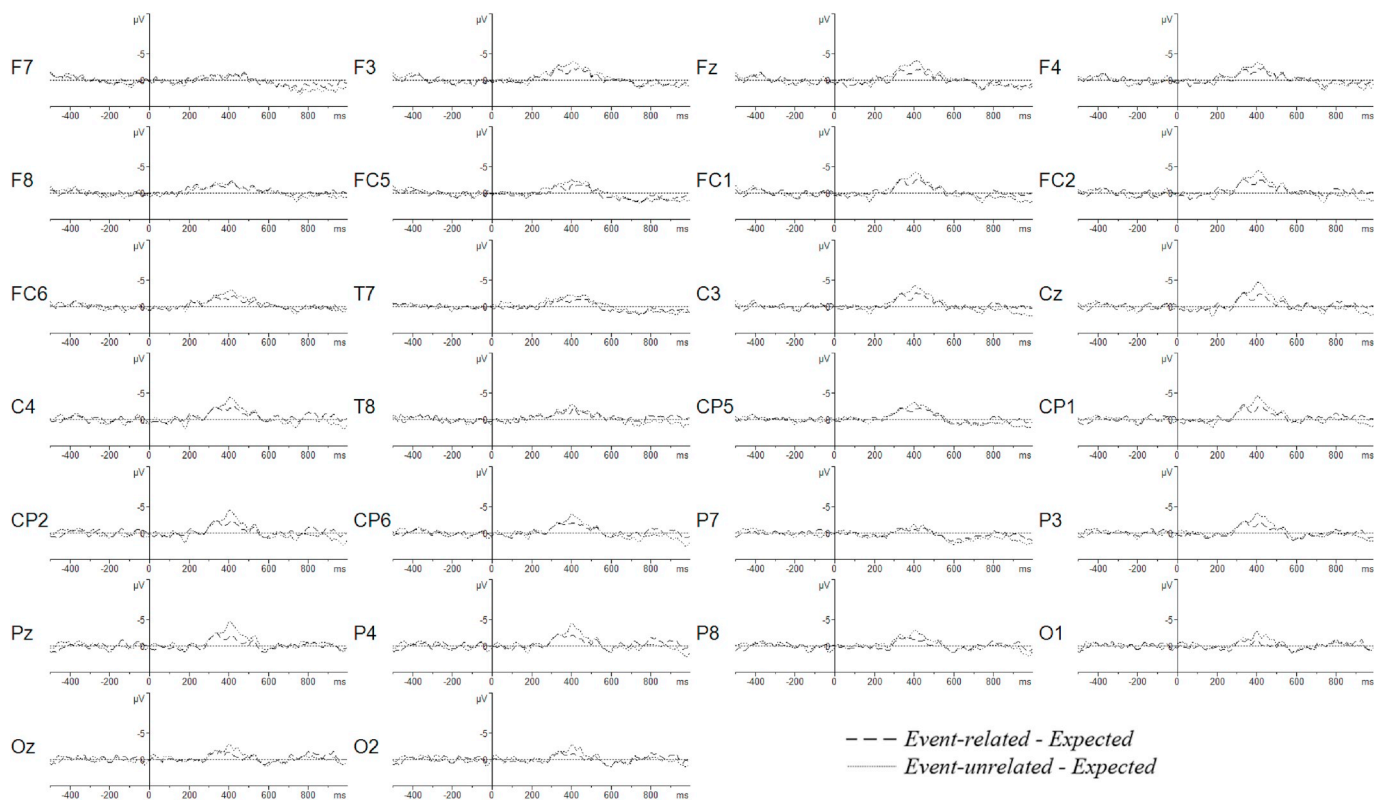


Fig. 7. Difference waves reflecting the size of N400 effects in the event-related and event-unrelated conditions in Experiment 2.

.011). The main effects and interactions were explored by post-hoc Bonferroni-corrected comparisons. The main effect of Difference echoes the previous analyses and highlights the consistent difference between event-related and event-unrelated conditions in both experiments ($p < .001$). The main effect of Experiment shows that N400 amplitudes were generally more negative in Experiment 1 than in Experiment 2. The Experiment-by-Difference interaction indicates that N400 amplitudes for event-related and event-unrelated conditions were more negative in Experiment 1 than in Experiment 2 (event-related: $p = .05$; event-unrelated: $p < .001$).

In contrast to Experiment 1, the plot of the grand averages in

Experiment 2 do not show any differences between the deflections elicited by each of the three conditions in the time windows preceding and following the N400 time window. This result is similar to Metusalem et al.'s Experiment 2, where participants read the passage-final sentences in isolation, and suggests that, in particular, the observed differences between the conditions in the late time window in Experiment 1 (and Metusalem et al.'s Experiment 1) were most likely connected to the presence of a coherent event discourse. In their General Discussion, Metusalem and colleagues speculate that these effects “might be linked in some way to the eliciting word’s status as related or unrelated to the described event [and that] late effects elicited by semantic violations

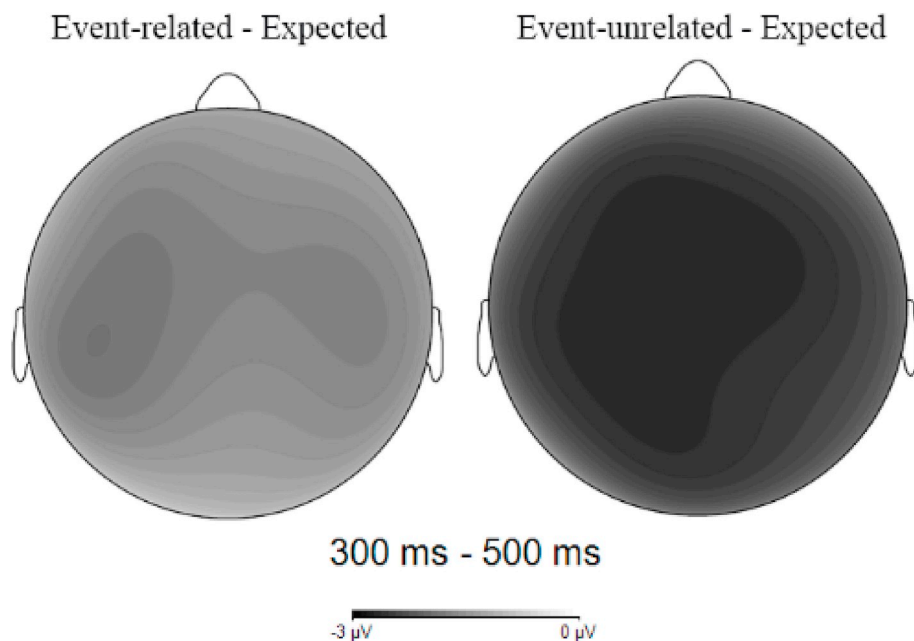


Fig. 8. Scalp topographies of the N400 effects in the event-related and event-unrelated conditions from Experiment 2.

[might] in some way [be] modulated by the eliciting word's degree of relation to the described event" (p. 560; cf. Kuperberg et al., 2006; Nieuwland and Van Berkum, 2005). As the current study was not designed to address this issue and our primary focus was on the N400 time window, our results do not allow us to make any conclusion about this speculation. Based on our manipulation, we are confident to conclude that simple associations between words did most likely not contribute to the late window effects. However, as far as their actual locus is concerned, future research is needed.

To summarize, using an associative priming manipulation, we obtained a similar three-way N400 split pattern as in Experiment 1: The N400 amplitude elicited by unexpected event-related target words lay between the ones elicited by the expected and unexpected event-unrelated target words.

7. General Discussion

In the current study, we tested the contribution of word associations to discourse comprehension. Dutch participants read short paragraphs consisting of three sentences. In Experiment 1, the three sentences formed a coherent discourse context. While the first two sentences established an event scenario, the paragraph-final sentence contained three kinds of target words: A highly expected word, an unexpected word that was related to the discourse context, or an unexpected word that was not related to the context. In Experiment 2, the three sentences did not form a coherent discourse context. The same paragraph-final sentences were preceded by two isolated sentences, each containing a prime word that was associatively related to the critical targets and part of the event-establishing sentences of Experiment 1. In Experiment 2, we thus reduced the likelihood of simulating a coherent event while maintaining parts of the lexical materials from Experiment 1, specifically still including the prime words that were associated with the targets.

The results of Experiment 1 replicate previous research (Metusalem et al., 2012, Experiment 1): Expected target words and unexpected event-unrelated target words elicited positive-going and strongly negative-going N400 amplitudes, respectively. The N400 amplitude elicited by the unexpected event-related target words was attenuated (e.g., less negative relative to the unexpected event-unrelated condition). These findings are consistent with the notion that comprehenders

simulate the described event and use their knowledge thereof to guide online language comprehension. Topographic analyses revealed that the N400 amplitudes elicited by both unexpected conditions were generally most negative over medial parietal-occipital electrodes, a region typically associated with the N400 and semantic processing (see Kutas and Federmeier, 2011; for discussion). Importantly, however, when the same targets followed sentences that did not establish a coherent event but included the associatively related prime words (Experiment 2), a similar N400 three-way split pattern was obtained as in Experiment 1. Consistent with our hypothesis, we interpret this finding as suggesting that associative priming led to the N400 amplitude reduction in the unexpected event-related condition.

This interpretation aligns well with conclusions drawn from previous experimental work arguing that event simulation and associations may contribute independently to discourse comprehension (Boudewyn et al., 2011, 2013; see also Kukona et al., 2011). On such an account, the effect in Experiment 1 would reflect the joint contributions of event simulation and associative priming; the effect in Experiment 2 would largely reflect the contribution of associative priming. Note that this proposal predicts that the difference between the event-related and the event-unrelated unexpected conditions should be larger in Experiment 1 than in Experiment 2, which was indeed borne out in our statistical comparison of Experiment 1 and 2 (see Results section of Experiment 2).

An alternative account for the data pattern observed in Experiment 2 is that participants (unconsciously) integrated the two unrelated introductory sentences and the target sentence as describing one connected event and that attempting to simulate this event led to the N400 amplitude reduction. However, as our rating study comparing event coherence across experiments suggests, the sentences in Experiment 2 were at best loosely connected, i.e. much less so than the events described in Experiment 1. When creating the sentences for the two prime words, we took several measures to make sure that the three sentences were maximally unrelated (e.g., agents, patients, and personal pronouns in the second and target sentence could not be linked back to the preceding sentence). We thus believe that integrating the three sentences into one event is an unlikely explanation for the results in Experiment 2.

However, a related possibility is that reading the prime words (e.g., 'baptized', 'church'), rather than activating the unexpected event-related target words, called up a whole event (e.g. baptism), which in

turn activated event-related objects and/or persons (e.g. 'priest'). Relevant prior research supporting such an account comes from an ERP study by Chwilla and Kolk (2005), who observed priming (reflected in a reduced N400 amplitude) in a word triplet paradigm. In their study, the combination of two primes pointed to an event to which the target word was related (e.g., 'director' and 'bribe' together primed 'dismissal'). Based on the current data, we cannot distinguish between word-word and word-event priming as the source of the N400 amplitude modulation in Experiment 2. That is, we cannot exclude the possibility that reading 'baptized' or 'church' led to the activation of the 'baptism event'. However, our manipulation was different from that of Chwilla and Kolk (2005) in that we placed each of the two prime words in successive sentences rather than presenting them as isolated words on the same screen. Thus, our prime words were separated by several intervening unrelated words. This was intended to reduce the likelihood of participants integrating both prime words into one discourse model, which would readily activate a given event.

How could lexical priming effects persist over such a long lag? It is often assumed that associative (word-word) priming effects decay quite rapidly. However, there are some examples in the literature that challenge this view. Based on the results from multiple behavioral and computational modelling experiments, Becker et al. (1997); see also Joordens and Becker, (1997) proposed that semantic priming may span word lists of up to eight items, provided that the semantic features of the prime words were fully activated or at least above a certain threshold. Given that our participants were instructed to read the two introductory sentences containing the prime words carefully, one may assume that the semantic features were fully activated and that semantic priming may have contributed to the activation of the event-unrelated targets in Experiment 2, at least in the word pairs bearing a semantic relationship—recall that we did not distinguish between purely semantic and purely associative relationships in the present materials. We speculate that in the pairs bearing a mere associative relationship, encountering the first prime word partially activated the event-related target and that encountering the second prime word some time later boosted its activation, preventing the rapid decay that is often reported in associative priming experiments (Neely, 1991). Future research could be conducted to investigate this conjecture experimentally.

As mentioned in the Introduction, Metusalem et al. (2012) have interpreted their findings as demonstrating that comprehenders use event simulations to generate predictions about upcoming referents during discourse reading. We subscribe to a view where event simulation and associations are two mechanisms contributing to prediction during comprehension (Huettig, 2015). However, a recent methodological debate (see Nieuwland et al., 2018) has led us to refrain from committing to an interpretation of the present data that attributes the N400 amplitude reduction in the event-related condition to conclusively reflect predictive processing. Specifically, as many scientists have pointed out previously, as the N400 component was measured on the target word itself (and not before), it is unclear whether its reduction reflects prediction or ease of integration.

A recent proposal holds that the (target word) N400 might reflect both prediction and integration processes, which show up as different spatio-temporal signatures (Nieuwland et al., 2019). The authors argue that semantic facilitation of predictable words might arise from a cascade of processes that (pre-)activate and integrate word meaning with context into a sentence- or discourse level meaning. A possibility for future research is to link this account to the current data and ask whether, and if so how strongly, associations and event simulation contribute to prediction and integration processes.

To conclude, we contrasted the contributions of event simulation and associations to discourse reading using event-related brain potentials. We observed that contextually unexpected target words that were related to an event description (which included the associatively related words) elicited a reduced N400 amplitude compared to contextually unexpected target words that were unrelated to the event. Crucially, a

similar N400 attenuation was observed when the likelihood of engaging in simulating a coherent event was substantially reduced, but the discourse context included the associatively related words. As the difference between event-related and event-unrelated conditions was larger when the sentences formed a coherent event than when they did not, our results suggest that associative priming alone cannot account for the N400 pattern observed in our Experiment 1. However, because part of the effect remained in Experiment 2, the findings fit best with the notion that during discourse reading both event simulation and simple associative mechanisms jointly contribute to the activation of words beyond contextually congruent sentence continuations.

Acknowledgements

We thank Annelies van Wijngaarden, Sophie Kirkels and Marloes Wensink for assistance in preparing the experiments, and Joost Rommers and Antje Strauß for valuable discussions on the manuscript.

References

- Arai, M., Keller, F., 2013. The use of verb-specific information for prediction in sentence processing. *Lang. Cognit. Process.* 28 (4), 525–560.
- Baggio, G., Hagoort, P., 2011. The balance between memory and unification in semantics: a dynamic account of the N400. *Lang. Cognit. Process.* 26, 1338–1367.
- Bar, M., 2007. The proactive brain: using analogies and associations to generate predictions. *Trends Cognit. Sci.* 11 (7), 280–289.
- Bar, M., 2009. The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364 (1521), 1235–1243.
- Barsalou, L.W., 2008. Cognitive and neural contributions to understanding the conceptual system. *Curr. Dir. Psychol. Sci.* 17 (2), 91–95.
- Becker, S., Moscovitch, M., Behrmann, M., Joordens, S., 1997. Long-term semantic priming: a computational account and empirical evidence. *J. Exp. Psychol. Learn. Mem. Cognit.* 23 (5), 1059.
- Beeman, M., 1993. Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. *Brain Lang.* 44 (1), 80–120.
- Bentin, S., 1987. Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain Lang.* 31 (2), 308–327.
- Bentin, S., McCarthy, G., Wood, C.C., 1985. Event-related potentials, lexical decision and semantic priming. *Electroencephalogr. Clin. Neurophysiol.* 60 (4), 343–355.
- Boudewyn, M.A., Gordon, P.C., Long, D., Polse, L., Swaab, T.Y., 2011. Does discourse congruence influence spoken language comprehension before lexical association? Evidence from event-related potentials. *Lang. Cognit. Process.* 27 (5), 698–733.
- Boudewyn, M.A., Long, D.L., Swaab, T.Y., 2013. Effects of working memory span on processing of lexical associations and congruence in spoken discourse. *Front. Psychol.* 4, 60.
- Brothers, T., Swaab, T.Y., Traxler, M.J., 2015. Effects of prediction and contextual support on lexical processing: prediction takes precedence. *Cognition* 136, 135–149.
- Brysbart, M., 2019. How many words do we read per minute? A review and meta-analysis of reading rate. *J. Mem. Lang.* 109, 104047.
- Camblin, C.C., Gordon, P.C., Swaab, T.Y., 2007. The interplay of discourse congruence and lexical association during sentence processing: evidence from ERPs and eye tracking. *J. Mem. Lang.* 56 (1), 103–128.
- Chen, E., Gibson, E., Wolf, P., 2005. Online syntactic storage costs in sentence comprehension. *J. Mem. Lang.* 52 (1), 144–169.
- Chiarello, C., Liu, S., Faust, M., 2001. Bihemispheric sensitivity to sentence anomaly. *Neuropsychologia* 39 (13), 1451–1463.
- Chwilla, D.J., Kolk, H.H., 2005. Accessing world knowledge: evidence from N400 and reaction time priming. *Cognit. Brain Res.* 25 (3), 589–606.
- Coulson, S., Federmeier, K.D., Van Petten, C., Kutas, M., 2005. Right hemisphere sensitivity to word- and sentence-level context: evidence from event-related brain potentials. *J. Exp. Psychol. Learn. Mem. Cognit.* 31 (1), 129–147.
- De Deyne, S., Navarro, D., Storms, G., 2013. Better explanations of lexical and semantic cognition using networks derived from continued rather than single word associations. *Behav. Res. Methods* 45 (2), 480–498.
- Elman, J.L., 2009. On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cognit. Sci.* 33 (4), 547–582.
- Elman, J.L., McRae, K., 2019. A model of event knowledge. *Psychol. Rev.* 126 (2), 252.
- Federmeier, K.D., Kutas, M., 1999. A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41 (4), 469–495.
- Federmeier, K.D., Kutas, M., 2002. Picture the difference: electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia* 40 (7), 730–747.
- Federmeier, K.D., McLennan, D.B., De Ochoa, E., Kutas, M., 2002. The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study. *Psychophysiology* 39 (2), 133–146.
- Ferreira, F., Chantavarin, S., 2018. Integration and prediction in language processing: a synthesis of old and new. *Curr. Dir. Psychol. Sci.* 27 (6), 443–448.
- Gratton, G., Coles, M.G.H., Donchin, E., 1983. A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484.

- Greenhouse, S.W., Geisser, S., 1959. On methods in the analysis of profile data. *Psychometrika* 24 (2), 95–112.
- 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., Steenbeek-Planting, E.G., Hagoort, P., 2007. The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Res.* 1146, 210–218.
- Huetting, F., 2015. Four central questions about prediction in language processing. *Brain Res.* 1626, 118–135.
- Huetting, F., Mani, N., 2016. Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience* 31 (1), 19–31.
- Huetting, F., Rommers, J., Meyer, A.S., 2011. Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* 137 (2), 151–171.
- Hutchison, K.A., 2003. Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychon. Bull. Rev.* 10 (4), 785–813.
- Joordens, S., Becker, S., 1997. The long and short of semantic priming effects in lexical decision. *J. Exp. Psychol. Learn. Mem. Cognit.* 23 (5), 1083.
- Kahneman, D., Tversky, A., 1973. On the psychology of prediction. *Psychological Review* 80 (4), 237.
- Keuleers, E., Brysbaert, M., New, B., 2010. SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behav. Res. Methods* 42 (3), 643–650.
- Kleiman, G.M., 1980. Sentence frame contexts and lexical decisions: sentence-acceptability and word-relatedness effects. *Mem. Cognit.* 8 (4), 336–344.
- Klem, G.H., Lüders, H.O., Jasper, H.H., Elger, C., 1999. The ten-twenty electrode system of the international federation of clinical neurophysiology. *Electroencephalogr. Clin. Neurophysiol.* (52), 3–6. Supplement 1999.
- Knoeferle, P., Crocker, M.W., Scheepers, C., Pickering, M.J., 2005. The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition* 95 (1), 95–127.
- Kukona, A., Fang, S.Y., Aicher, K.A., Chen, H., Magnuson, J.S., 2011. The time course of anticipatory constraint integration. *Cognition* 119 (1), 23–42.
- Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146 (1), 23–49.
- Kuperberg, G.R., Caplan, D., Sitnikova, T., Eddy, M., Holcomb, P.J., 2006. Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Lang. Cognit. Process.* 21 (5), 489–530.
- Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647.
- Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207 (4427), 203–205.
- Landauer, T.K., Dumais, S.T., 1997. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104 (2), 211–240.
- Laszlo, S., Stites, M.C., Federmeier, K.D., 2012. Won't get fooled again: An event-related potential study of task and repetition effects on the semantic processing of items without semantics. *Language Cogn. Process.* 27, 257–274.
- Ledoux, K., Camblin, C.C., Swaab, T.Y., Gordon, P.C., 2006. Reading words in discourse: the modulation of lexical priming effects by message-level context. *Behav. Cognit. Neurosci.* 5 (3), 107–127.
- Mandler, G., 1984. *Mind and Body: Psychology of Emotion and Stress*. WW Norton, New York.
- Mantegna, F., Hintz, F., Ostarek, M., Alday, P.M., Huetting, F., 2019. Distinguishing Integration and Prediction Accounts of ERP N400 Modulations in Language Processing through Experimental Design. Advance online publication. <https://doi.org/10.1016/j.neuropsychologia.2019.107199>. *Neuropsychologia*.
- McRae, K., Matsuki, K., 2009. People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass* 3 (6), 1417–1429.
- Metusalem, R., Kutas, M., Urbach, T.P., Hare, M., McRae, K., Elman, J.L., 2012. Generalized event knowledge activation during online sentence comprehension. *J. Mem. Lang.* 66 (4), 545–567.
- Minsky, M., 1974. A framework for representing knowledge. In: Winston, P. (Ed.), Reprinted in *the Psychology of Computer Vision*. McGraw-Hill, 1975.
- Neely, J.H., 1991. Semantic priming effects in visual word recognition: a selective review of current findings and theories. In: Besner, D., Humphreys, G. (Eds.), *Basic Processes in Reading: Visual Word Recognition*. Erlbaum, Hillsdale (New Jersey, US), pp. 264–336.
- Nelson, D.L., McEvoy, C.L., Schreiber, T.A., 2004. The University of South Florida free association, rhyme, and word fragment norms. *Behav. Res. Methods* 36 (3), 402–407.
- Nieuwland, M.S., Van Berkum, J.J.A., 2005. Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognit. Brain Res.* 24 (3), 691–701.
- Nieuwland, M.S., et al., 2019. Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials. *Phil. Trans. R. Soc. B* 375, 20180522. <https://doi.org/10.1098/rstb.2018.0522>.
- Nieuwland, M.S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsturn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D.J., Rousselet, G., Ferguson, H.J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E.M., Donaldson, D.I., 2 others, 2018. Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife* 7, e33468.
- Noyes, J.M., Garland, K.J., 2008. Computer- vs. paper-based tasks: are they equivalent? *Ergonomics* 51 (9), 1352–1375.
- Otten, M., van Berkum, J.J.A., 2007. What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Res.* 1153, 166–177.
- Otten, M., Van Berkum, J.J.A., 2008. Discourse-based word anticipation during language processing: prediction or priming? *Discourse Process* 45 (6), 464–496.
- Paczynski, M., Kuperberg, G.R., 2012. Multiple influences of semantic memory on sentence processing: distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *J. Mem. Lang.* 67 (4), 426–448.
- Pickering, M.J., Gambi, C., 2018. Predicting while comprehending language: a theory and review. *Psychol. Bull.* 144 (10), 1002–1044.
- Pickering, M.J., Garrod, S., 2013. An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36 (4), 329–347.
- Rommers, J., Meyer, A.S., Praamstra, P., Huetting, F., 2013. The contents of predictions in sentence comprehension: activation of the shape of objects before they are referred to. *Neuropsychologia* 51 (3), 437–447.
- Rugg, M.D., 1987. Dissociation of semantic priming, word and non-word repetition effects by event-related potentials. *The Quarterly J. Experimental Psychology Section A* 39 (1), 123–148.
- Rumelhart, D.E., 1980. On evaluating story grammars. *Cognit. Sci.* 4 (3), 313–316.
- Sanford, A.J., Garrod, S.C., 1998. The role of scenario mapping in text comprehension. *Discourse Process* 26 (2–3), 159–190.
- Sanford, A.J., Leuthold, H., Bohan, J., Sanford, A.J.S., 2011. Anomalies at the borderline of awareness: an ERP study. *J. Cognit. Neurosci.* 23, 514–523.
- Schank, R., Abelson, R., 1977. *Scripts, Plans, Goals and Understanding*. Erlbaum, Hillsdale, NJ.
- Schwanenflugel, P.J., Shoben, E.J., 1985. The influence of sentence constraint on the scope of facilitation for upcoming words. *J. Mem. Lang.* 24 (2), 232–252.
- Staub, A., Clifton Jr., C., 2006. Syntactic prediction in language comprehension: evidence from either...or. *J. Experimental Psychology.: Learn., Mem., and Cognit.* 32 (2), 425.
- Taylor, W.L., 1953. "Cloze procedure": a new tool for measuring readability. *Journalism Bulletin* 30 (4), 415–433.
- Van Berkum, J.J.A., Hagoort, P., Brown, C.M., 1999. Semantic integration in sentences and discourse: evidence from the N400. *J. Cognit. Neurosci.* 11 (6), 657–671.
- Van Berkum, J.J., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cognit.* 31 (3), 443.
- Van Petten, C., 1993. A comparison of lexical and sentence-level context effects in event-related potentials. *Lang. Cognit. Process.* 8 (4), 485–531.
- Van Petten, C., 2014. Examining the N400 semantic context effect item-by-item: relationship to corpus-based measures of word co-occurrence. *Int. J. Psychophysiol.* 94 (3), 407–419.
- Wicha, N.Y., Bates, E.A., Moreno, E.M., Kutas, M., 2003. Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neurosci. Lett.* 346 (3), 165–168.
- Wicha, N.Y., Moreno, E.M., Kutas, M., 2004. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *J. Cognit. Neurosci.* 16 (7), 1272–1288.
- Zacks, J.M., Tversky, B., 2001. Event structure in perception and conception. *Psychol. Bull.* 127 (1), 3.
- Ziefle, M., 1998. Effects of display resolution on visual performance. *Hum. Factors: The J. Human Factors and Ergonomics Soc.* 40 (4), 554–568.