Predictive processing and pragmatic inferencing

How does the brain deal with Dutch *eigenlijk*?

Marlou Rasenberg

Predictive processing and pragmatic inferencing

How does the brain deal with Dutch *eigenlijk*?

Master Thesis

Marlou Rasenberg

Supervisor:

Dr. Geertje van Bergen

Max Planck Institute for Psycholinguistics

Nijmegen, The Netherlands

Second reader:

Dr. Joost Rommers

Donders Institute for Brain, Cognition and Behaviour

Nijmegen, The Netherlands

Student number: 4597877

Programme: Research Master in Language and Communication

Faculty of Arts, Radboud University Nijmegen

Submission date: October 3, 2017

Acknowledgements

This thesis is the result of my internship within the Neurobiology of Language Department of the Max Planck Institute for Psycholinguistics. I would like to thank all NBL members for the pleasant teamwork and valuable feedback on the project. Over the past year I have gained a lot of experience and knowledge, and I am grateful to have gotten the opportunity to work in such an inspiring research environment.

Especially, I would like to thank my supervisor Geertje van Bergen for her great support and for guiding me through every step of the process. I am grateful for her enthusiasm for the subject and our many interesting discussions, and for helping me find my way in the academic world. Thank you for always making time for me, even on days when that must have been difficult. Many thanks also go to Joost Rommers for his expertise and help with the design and data analysis, and his willingness to be the second reader for this thesis.

Abstract

The Dutch discourse particle *eigenlijk* signals a mismatch between the utterance in which it occurs and the speaker's estimation of the hearer's expectation (Van Bergen, Van Gijn, Hogeweg, & Lestrade, 2011). The present study set out to investigate how this expectation-managing device affects predictive processing during incremental language comprehension. We used written mini-conversations with a context sentence and a question which evoked expectations about upcoming information. The answer which followed contained either *eigenlijk* or a neutral adverb, and a critical word which was either coherent or incoherent with respect to the prior discourse context. We first conducted an off-line discourse completion test, for which we truncated the answers. The test revealed that answers with *eigenlijk* evoked more contrastive continuations of the conversation compared to the neutral condition. In an ERP experiment, we asked whether and how *eigenlijk* affects the pre-activation and integration of coherent and incoherent discourse continuations. In line with previous studies, we found more negative amplitudes in the N400 time window for incoherent versus coherent critical words. However, this N400 effect was shown to be unaffected by the presence of *eigenlijk*. There were no significant differences between conditions with respect to late frontal positivities. For late posterior positivites, on the other hand, we found more positive amplitudes for critical words which were preceded by *eigenlijk* versus those preceded by a neutral adverb. We argued that this P600 effect could reflect pragmatic inferencing to uncover the speaker's intended meaning. Together these results seem to indicate that even though *eigenlijk* does not facilitate lexical-semantic processing, it does induce the updating of discourse representations in terms of socio-pragmatic inferences.

Table of contents

1. Introduction

Language comprehension is a universal skill. Whether it is a deep conversation or just everyday small talk; we routinely engage in interactions with others and are able to rapidly and automatically comprehend speech. Unlike other skills, language comprehension is something that comes naturally to (most of) us, yet upon closer inspection this seemingly effortless process is in fact very complex. Not only do we have to process about 150 words a minute, at the same time we have to make sense of what the speaker is trying to convey with these words *and* plan our next turn. To arrive at fast and accurate comprehension, our language system adopts an efficient strategy: rather than waiting for the words to be processed bottom-up, we anticipate what will come up next. We form these predictions based on our *discourse model*; i.e. our mental representation of the current communicative situation. This representation includes what has been said by whom previously, our real-world knowledge and, crucially, *common ground*. Common ground can be defined as the set of knowledge and beliefs that interlocutors share and mutually belief they share (Clark, 1996).

This general idea – that comprehenders anticipate what will come up next based on a given discourse context – has been confirmed by numerous psycholinguistic studies. In well-controlled experimental settings, researchers have measured anticipatory eye movements to objects or entities (with the Visual World Paradigm, e.g., Altmann & Kamide, 1999) and investigated event-related brain potentials while subjects process (un)expected linguistic information using EEG (e.g., Federmeier & Kutas, 1999; Kutas & Hillyard, 1984). Together, these studies have shown that comprehenders predict upcoming information, which facilitates the (semantic) integration of upcoming words. However, we remain humans, and humans make mistakes – or in this case, incorrect predictions. These disconfirmed predictions could come with a price in the form of extra processing costs (e.g., Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007, see Kuperberg & Jaeger, 2016 and Petten & Luka, 2012 for reviews).

This study deals with the type of incorrect predictions that can arise due to asymmetries in the knowledge status of interlocutors and the linguistic devices that we have at our disposal to modulate these expectations. We will focus on one such linguistic expectation-managing device: the Dutch discourse particle *eigenlijk*. Consider the following (constructed) dialogue:

(1) A *Je trekt toch wel iets netjes aan?*
You are going to dress up, aren't you?

B *Ik was <u>net</u> van plan een pak aan te trekken.*
I was just about to put on a suit.

B' *Ik was <u>eigenlijk</u> van plan een spijkerbroek aan te trekken.*
I was [actually] about to put on jeans.

The suggestive question by speaker A indicates that he or she expects speaker B to dress up, and thus to confirm his or her question. The reply can either match (B) or mismatch (B') this expectation. The unexpected word 'jeans' in B' is preceded by the particle *eigenlijk* [≈ actually, in fact], which is theoretically assumed to mark the asymmetry in belief states and the resulting unexpectedness of the utterance (Van Bergen, Gijn, Hogeweg, & Lestrade, 2011, p. 3885).

In this study we investigate how the expectation-managing particle *eigenlijk* affects on-line predictive processing in a conversational context. Specifically, we are interested in the possible facilitating effect of *eigenlijk* when it precedes unexpected information. Accordingly, the research question addressed in this thesis is:

RQ: What is the effect of eigenlijk *on the processing of (un)expected linguistic information during incremental language comprehension?*

We address this question by means of two experiments. First, we explore how the particle *eigenlijk* modulates predictions about upcoming information in an off-line discourse completion test. Subsequently, we use the results and stimuli of this pretest in an ERP experiment in which we investigate the effect of *eigenlijk* on on-line sentence processing. We thus adopt an interdisciplinary approach, combining insights and methods from both theoretical linguistics and cognitive neuroscience, whereby we are the first to conduct an ERP study on expectation-managing discourse particles. The results of this project can contribute to the research field in two ways: on the one hand this study can provide empirical evidence for the theoretically assumed functions of *eigenlijk,* while on the other hand discourse particles are useful tools to investigate the facilitating effect of discourse-based predictions on language processing, and can help us to understand the brain signals we pick up.

This thesis is structured as follows. First, we provide background information on theoretical accounts of discourse particles, as well as information on predictive processing and findings from previous ERP studies. After presenting the research hypotheses, we proceed to the methods and results of the discourse completion test and ERP study. Finally, we interpret the results and provide conclusions.

2. Background

2.1 Discourse particle eigenlijk

In the introduction we stated that we make inferences about upcoming information in a conversation based on our discourse model. This includes speaker knowledge, for which we introduced the notion *common ground*. Common ground is the set of knowledge which is shared and believed to be shared between interlocutors (Clark, 1996). When information is only known by one of the interlocutors, we speak of *privileged ground*. Note, however, that the knowledge status of interlocutors is not static: “Because discourse involves the exchange of information, knowledge and meta-knowledge are constantly in flux, as are degrees of certainty and salience” (Schiffrin, 1988, p. 28).

In order to efficiently manage this interactive process of updating information states, we have a group of linguistic devices at our disposal: *discourse markers*. Discourse markers (e.g. *well, so, anyway*; hereafter referred to as DMs) are linguistic expressions which signal a relation between the utterance in which they occur and the preceding discourse (Fraser, 1999). The type of discourse relations which can be marked with DMs have been categorized in different ways. Fraser (1999) distinguishes between relations at the propositional, epistemic or speech act level. Maschler and Schiffrin (Maschler & Schiffrin, 2015) discern four domains: expressive, social, cognitive and textual. In general, it is important to note that the use of a DM is not restricted to one particular domain: DMs are multifunctional and their interpretation depends on the context in which they occur. Even though DMs are syntactically optional, they are highly frequent in everyday conversations. They help the hearer or reader, as they “constrain or guide the interpretation process” (Aijmer & Simon-Vandenbergen, 2004, p. 1784) and as such they are “communicatively obligatory” (Diewald, 2010, p. 32).

Let us now consider one such DM in more detail: the Dutch discourse particle *eigenlijk*. *Eigenlijk* – like all DMs – has multiple functions, but relevant for this study is that it expresses *interpersonal* or *intersubjective* meanings: it marks a mismatch between the interlocutors’ discourse representations (Van Bergen et al., 2011). More specifically, it signals a contrast between what the speaker knows (*privileged ground*) and what the speaker believes the interlocutor to know, and as such it functions as a common ground managing device. *Eigenlijk* thus requires “a theory of mind on the part of speaker, in that it presupposes a number of implicit assumptions about the state of mind of the hearer” (Van Bergen et al., 2011, p. 3891).

Apart from this function at the epistemic level, *eigenlijk* also serves a socio-pragmatic goal: by putting him- or herself in the hearer’s shoes and using *eigenlijk,* the speaker indicates that the expectations of the hearer are incorrect, but legitimate given the hearer’s discourse representation. As such, *eigenlijk* serves a face-saving strategy (Van Bergen, submitted; Van Bergen & Hogeweg, under review).

Thus, from a production perspective, it is well-defined when and why speakers use the particle *eigenlijk*: in sum, "*eigenlijk* expresses the unexpectedness of the utterance in which it appears given the speaker's estimation of the hearer's belief" (Van Bergen et al., 2011, p. 3885). In this study we will focus on the question how *eigenlijk* affects discourse comprehension, and more specifically, whether it can modulate expectations about upcoming words. Important to note about the use of *eigenlijk* is that it usually (though not necessarily) occurs in sentence-initial or sentence-medial position, and as such it *precedes* the information which it marks as unexpected for the hearer (see again example (1)). Therefore, *eigenlijk* could be characterized as a cue or warning that unexpected information will follow – a cue which the hearer could use to modulate predictions during incremental language processing. To further substantiate this hypothesis, we will need information on when, how and why we engage in predictive processing. We turn to these questions in the upcoming sections.

2.2 Predictive processing

2.2.1 Predictive pre-activation

Nowadays there is a general consensus that language processing is predictive. However, an issue that has lingered for some time is the *integration* vs. *prediction* debate; integrative accounts posture that context information asserts its facilitating effect only *after* word recognition – easing the integration of the stimulus into the discourse model –, whereas predictive accounts argue for *pre-activation* of information prior to encountering the word (see Federmeier, 2007; Petten & Luka, 2012 for reviews).

Recently the latter – strong prediction – view is winning ground, and it is generally acknowledged that comprehenders make predictions based on a discourse context in a graded, probabilistic fashion. ERP studies have yielded evidence for the pre-activation of semantic features (Federmeier & Kutas, 1999), as well as morpho-syntactic features (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004). Prediction could potentially even involve the pre-activation of the phonological form of words, as has been shown by an influential study by DeLong, Urbach and Kutas (2005) – but see Nieuwland et al. (2017) and commentaries (DeLong, Urbach & Kutas 2017; Yan, Kuperberg, & Jaeger, 2017) for counter-evidence and discussion. All in all, Kuperberg and Jaeger (2016) nicely sum up the existing evidence and conclude that "at least under some circumstances, higher-level information within our internal representations of context can lead to the pre-activation of incoming information at multiple lower level representations" (p. 42).

Other accounts even include the possibility that we *pre-update* our discourse model with the predicted information prior to encountering the linguistic input itself – also described in terms of predictive *commitment* (Kuperberg & Jaeger, 2016; Lau, Holcomb, & Kuperberg, 2013). However, such a commitment could be costly in the case of a disconfirmed prediction. It has been suggested that critical words that violate a strong lexical prediction can result in extra processing costs (see DeLong, Troyer, & Kutas, 2014 and Petten & Luka, 2012 for reviews). In the next section we turn to the ERP components that are used as indicators for the accompanying benefits and costs of predictive processing.

2.2.2 ERP components

Psycholinguists have studied predictive processing in various ERP studies, and a component that has been studied extensively is the N400. The N400 is a negativity peaking around 400 ms after stimulus onset, and is taken to reflect the ease of semantic processing, which is modulated by the presence of supportive context information. In a groundbreaking study, Kutas and Hillyard (Marta Kutas & Hillyard, 1984) found that unexpected words ("He liked lemon and sugar in his *coffee*") elicit larger negative-going deflections between 300-500 ms after word onset than expected words ("He liked lemon and sugar in his *tea"*). This study also showed that the effect of context on the N400 amplitude for a word is graded, as amplitude reductions are correlated with the word's *cloze probability.* This is a measure derived from so-called *cloze tests* and is calculated as the probability that subjects complete a context with a particular word.

Thus words that are expected to follow a certain context (as indicated by their cloze probability) are more easily processed (as indicated by a smaller N400). To determine whether this is due to pre-activation of upcoming information, Federmeier and Kutas (1999) further refined the paradigm by constructing contexts with critical words that are either from the same semantic category (*palms* and *pines*) or from different categories (*palms* and *tulips*). An example of a context is: "They wanted to make the hotel look more like a tropical resort. So, along the driveway, the planted rows of…". The authors hypothesized that if comprehenders predict the upcoming word *palms* and pre-activate its semantic features (e.g. "tree", "green", "tropical"), this would in turn result in facilitated processing for words with overlap in semantic features (*pines),* but not for words from a different semantic category (*tulips*). The results confirmed this hypothesis, as the N400 was smallest for the expected continuation *palms,* largest for the semantically unrelated and unexpected *tulips,* and in-between for the related but unexpected *pines*. Crucially, the difference between the latter two conditions disappeared in weak-constraining contexts (i.e., contexts that do not evoke a particular prediction), which shows that the facilitation was due to a strong prediction and pre-activation of semantic features.

Taken together, we can characterize the N400 as reflecting semantic memory access, and as such "the N400 response to a given input can be used as a tool to assess semantic memory states, with the amount of N400 reduction (relative to a control condition) revealing how much of the information normally elicited by that stimulus is already active" (Marta Kutas & Federmeier, 2011, p. 23). Important to note, however, is that the N400 can only indicate the benefits of a supportive context for processing – it does not reflect *prediction error* in the case of disconfirmed predictions. This has been shown in a study by Federmeier et al. (2007), in which they contrasted expected with unexpected endings of contexts that are either high-constraining (e.g., "The child was born with a rare *disease/gift*") or weak-constraining (e.g. "Mary went into her room to look at her *clothes/gift*"). They found that processing was facilitated for words with the best fit in the context (*disease* and *clothes*), with the largest facilitation for expected words in a strong-constraining context (*disease* is strongly predicted and thus evoked the smallest N400). However, there is no difference between the unexpected word (*gift)* across high- and weak-constraining contexts. This shows that prediction error (predicting *disease* but encountering *gift)* is not reflected by the N400.

Interestingly, in the same study, Federmeier et al. (2007) found differences across those two conditions in a later time window: unexpected words that violate a strong prediction evoked a frontally distributed positivity between 500-800 ms after word onset, in comparison to the unexpected word in a weak-constraining context which did not evoke such a waveform. Other studies have revealed similar patterns in response to unexpected continuations of high-constraining contexts (Delong, Urbach, Groppe, & Kutas, 2011; DeLong, Quante, & Kutas, 2014; Kutas, 1993; Moreno, Federmeier, & Kutas, 2002; Thornhill & Van Petten, 2012), and these ERP patterns have been termed frontal post-N400 positivity (PNP) or anterior late positivity (see DeLong et al., 2014; Van Petten & Luka, 2012, for reviews). It remains up to further research to unravel the exact functionality underlying this ERP component; it has been suggested that it reflects the inhibition or revision of a strong but disconfirmed prediction (Federmeier, 2007), but it could also be due to other cognitive processes (DeLong et al., 2014). In the most minimal sense, these late frontal positivities are taken to reflect "recruitment of additional processing when linguistic information is contextually pre-activated but not encountered" (DeLong et al., 2014, p. 640).

The frontal positivity component is thus a relatively recent discovery, and findings do not yet completely converge. Nevertheless it should be distinguished from other late positivity effects, such as the P600. The P600 is a posteriorly distributed positivity starting about 500 ms after word onset. Initially this component was taken to reflect syntactic parsing difficulties, but it has also been found for words which violate semantic constraints, for example with respect to thematic roles (e.g., Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006). Subsequently, it has been proposed that the P600 reflects a "combinatorial analysis" of the output of multiple processing streams (Kuperberg, 2007) or general reanalysis following a strong violation (Van de Meerendonk, Kolk, Chwilla, & Vissers, 2009). However, more recently, the P600 has also been reported in studies on pragmatic phenomena, such as jokes (e.g., Coulson & Lovett, 2004), indirect requests (e.g., Coulson & Lovett, 2010) and irony (e.g., Regel, Gunter, & Friederici, 2011; Spotorno, Cheylus, Henst, & Noveck, 2013). Based on the results from studies on syntactic, semantic as well as pragmatic phenomena, Brouwer, Fitz and Hoeks (2012) have put forward a more holistic account of the underlying functionality of P600 effects. According to Brouwer et al., lexical-semantic memory retrieval takes place during the N400 time-window, and the activated information is subsequently integrated into the discourse representation during the P600 time-window. They conclude that "the P600 can best be understood in terms of the construction, reorganization, or updating of a mental representation of what is being communicated in a sentence or story" (p. 140).

To sum up, there are thus two positive-going waveforms with an overlapping time window (both roughly 600-900 ms after word onset): the frontal positivity and the P600. The two components differ in their scalp distributions (frontal versus posterior), as well as the type of stimuli that elicit these effects. First of all, frontal positivities reflect the disconfirmation of *lexical* predictions, rather than predictions at the semantic or conceptual level, as shown by Thornhill and Van Petten (2012). Second, in one of the few studies that have directly compared the two components, DeLong, Quante and Kutas (2014) have shown that frontal positivity effects are only found for words that are unexpected, but still *plausible* given the prior context. As such, "these words have the potential to immediately be made sense of in their contexts, without further input" (p. 161). This is in contrast to contexts followed by semantically anomalous words, which evoked a posterior P600 effect. The authors have thereby shown that "anomalous and merely unlikely written sentence continuations are processed in qualitatively different ways" (p. 161).

2.3 Reversing predictions

Comprehenders are thus able to use higher level information to predict upcoming information, which facilitates the rapid retrieval and integration of the upcoming words within their discourse representation. Initially this has been shown at the sentence-level (e.g., "He liked lemon and sugar in his *coffee*", Kutas & Hillyard, 1984), but these findings generalize to the discourse-level, as information from prior sentences (rather than the carrier sentence) can facilitate the processing of upcoming words (e.g., Van Berkum, Zwitserlood, Hagoort, & Brown, 2003). Such effects of the wider context have been shown to be very strong; for example, hearing the sentence "The peanut was in love" would normally result in processing difficulties (as indicated by the N400), but not when this sentence is presented in a cartoon-like discourse context (Nieuwland & Berkum, 2006).

Apart from this linguistic context information, comprehenders use more general knowledge to process information. This can be real-world knowledge based on individual's experiences, which is stored in long-term memory (e.g., knowledge about trains in The Netherlands help to anticipate information in sentences such as "the Dutch trains are *yellow*", Hagoort, Hald, & Bastiaansen, 2004). But we can also make inferences about a speaker based on his or her voice, and use this information to anticipate upcoming speech (e.g., when listening to someone with an upper-class accent, hearing "I have a big *tattoo* on my back" elicits an N400 effect, Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008).

However, the crucial question for this study is whether comprehenders can, in addition to the discourse context and their real-world knowledge, use the particle *eigenlijk* as a cue for predictive processing. This would require comprehenders to incrementally integrate the higher-level information encoded in specific linguistic expressions in their current discourse representation. This has been investigated for the intersubjective discourse particles *eigenlijk* and *inderdaad* in a recent eye-tracking study by Van Bergen (submitted), adopting the Visual World Paradigm (Altmann & Kamide, 1999). Participants listened to short conversations, consisting of a high-constraining context and question, followed by a response containing *inderdaad, eigenlijk* or a control adverb. While listening, participants watched a screen with four pictures, which figured as potential referents for the upcoming information. In the control condition participants showed anticipatory looks to the referent in line with the expected conversational continuation. When participants encountered *eigenlijk,* the looks to this expected referent immediately decreased, instead showing increased attention to a plausible alternative referent. These results thus show that listeners immediately integrate the information encoded in *eigenlijk* to modulate predictions about upcoming referents. This is in line with other eye-tracking studies which have shown influences of specific linguistic cues on anticipatory processing, such as disfluencies (Bosker, Quené, Sanders, & de Jong, 2014; Corley & Hartsuiker, 2003), causal connectives (Canestrelli & Mak, 2013) and Dutch *er* [there] (Grondelaers, Speelman, Drieghe, Brysbaert, & Geeraerts, 2009).

Thus, eye-tracking studies have shown the influence of specific linguistic cues on anticipatory gaze patterns prior to encountering bottom-up information, indicating predictive processing. However, these studies do not provide a direct, on-line measure of the ease of lexical-semantic retrieval and integration when the (un)predicted information is encountered. For this we have to turn to ERP-studies. To our knowledge, there are no such studies investigating the effect of discourse particles on semantic processing, but psycholinguists did study other specific linguistic expressions. For example, it has been found that verbal disfluencies facilitate the processing of unexpected information (Corley, MacGregor, & Donaldson, 2007), that negation is incrementally integrated to interpret sentences (Nieuwland & Kuperberg, 2008) and that 'if'-clauses are used to rapidly set up a counterfactual discourse model to which incoming words are linked (Nieuwland & Martin, 2012).

Furthermore, a study by Xiang and Kuperberg (2015) is of particular interest, as they have investigated the concessive connective *even so,* which signals that the upcoming information will contrast with the default expectation given a certain discourse context. As such, it is said to "pragmatically constrain" discourse comprehension. They measured ERPs to critical words which were either coherent or incoherent with respect to the discourse, and occurred with or without *even so,* e.g., "Elizabeth had a history exam on Monday. She took the test and aced/failed. (Even so,) she went home and *celebrated* wildly". The results confirmed the hypothesis that *even so* is used as a cue in incremental language comprehension, as it reverses and enhances semantic predictions. They found that the semantic processing of incoherent scenarios was facilitated when preceded by *even so*, as indicated by an attenuated N400 effect. However, the pragmatically odd use of *even so* (in combination with a discourse coherent scenario) elicited prolonged neural costs, as reflected by an increased P600 relative to the control condition. The authors argue that *even so* induced comprehenders to generate a strong alternative prediction, and that the P600 effect reflects an attempt to integrate the encountered information to construct a new discourse representation.

An important remark with respect to the ERP studies mentioned so far is that they investigated the influence of certain linguistic expressions on predictive processing in single sentences or short texts, and not much is known about the extent to which these findings would generalize to dialogue situations. Typical for the use of *eigenlijk* to express intersubjectivity, however, is its occurrence in interactions between interlocutors. The fact that there is only a limited amount of studies using (spoken) conversational stimuli is perhaps not surprising when one thinks about the challenges involved in creating such stimuli, which have to be natural and experimentally controlled at the same time. In fact, it has been said that "research on dialogue using EEG is in its infancy" (Bögels, Kendrick, & Levinson, 2015, p. 3). Nonetheless, there are a few EEG studies which indicate that comprehenders anticipate the timing and content of turns in conversations (Bögels et al., 2015; Magyari, Bastiaansen, de Ruiter, & Levinson, 2014), and can rapidly recognize speech acts (Gisladottir, Chwilla, & Levinson, 2015). This is promising, because in the present study the question is not only whether *eigenlijk* influences predictive processing, but more specifically we are asking whether comprehenders can predict upcoming information based on the discourse context, to then subsequently modulate these predictions as a result of the rapid integration of the complex intersubjective information encoded in *eigenlijk*, and do so over the turns of two different speakers (cf. Bögels, Kendrick, & Levinson, 2015).

To sum up, previous studies have shown that comprehenders can rapidly integrate information encoded in certain linguistic cues to modulate (or completely reverse) predictions about upcoming information. Furthermore, some pioneering EEG studies have shown that comprehenders also engage in predictive processing while listening to dialogue. In the present study we will investigate whether the Dutch discourse particle *eigenlijk* is used as a cue for predictive processing in a conversational setting, for which the hypotheses will be laid out in the next section.

3. Present study

We set out to examine whether and how the intersubjective discourse particle *eigenlijk* affects the pre-activation and integration of (un)expected information during incremental language comprehension in a conversational setting. To investigate this, we set up an ERP study in which participants read short conversations, consisting of a context sentence and a question-answer pair (cf. Van Bergen, submitted). The context and question are (medium to high) constraining, and the answer contains a critical word (CW) to which the EEG-signals are time-locked. Four different versions of the answer were created by crossing two factors: Particle (the presence of the discourse particle *eigenlijk* or a neutral adverb: Eigenlijk or Neutral) and Coherence (the coherence between the CW and the expectation raised by the context and question: Coherent or Incoherent).

Based on the reviewed literature, we have several predictions. First of all, we expect comprehenders to predict upcoming information based on the discourse context, and to pre-activate semantic features. As such, the effect of Coherence will be similar to previous studies, and we thus hypothesize the following:

H1: *If comprehenders predict upcoming information and pre-activate semantic features, the N400 for Coherent CWs should be smaller than for Incoherent CWs.*

H2: *If comprehenders generate strong, specific lexical predictions, there will be a larger frontal positivity for Incoherent CWs (due to a prediction error) compared to Coherent CWs.*

With respect to *eigenlijk,* the eye-tracking study by Van Bergen (submitted) has already shown that comprehenders can immediately integrate the high-level intersubjective information encoded in *eigenlijk* to modulate predictions about upcoming input. Furthermore, based on the reviewed literature, we argue that specific linguistic cues can facilitate the semantic processing of unexpected or incoherent information, as reflected in ERP components such as the N400 (e.g., Xiang & Kuperberg, 2015). We could extend this to *eigenlijk* in two ways: either comprehenders only become less certain about their initial prediction (in other words: they retract their predictive commitment) or they start anticipating (and pre-activating semantic features of) a contrastive continuation of the utterance. We therefore hypothesize the following effects of *eigenlijk* for the processing of Incoherent CWs:

H3a: *If* eigenlijk *only affects the strength of initial predictions, there will be no difference in N400 amplitude between the Eigenlijk-Incoherent and Neutral-Incoherent condition.*

H3b: *If comprehenders, upon encountering* eigenlijk, *revise their initial prediction and start pre-activating contrastive semantic features, there will be an attenuated N400 for the Eigenlijk-Incoherent condition compared to the Neutral-Incoherent condition.*

H4: *If comprehenders generate strong, lexical predictions (cf. H2) and* eigenlijk *affects the strength of these initial predictions (cf. H3a/b), we expect an attenuated frontal positivity for the Eigenlijk-Incoherent condition compared to the Neutral-Incoherent condtion.*

For CWs which are coherent with respect to the overall discourse, but preceded by *eigenlijk,* it is not straightforward whether the semantic processing will be affected. If *eigenlijk* evokes comprehenders to become less certain about their initial prediction, the semantic features of the coherent CWs might still be pre-activated. Alternatively, if comprehenders completely revise their predictions, there might be cross-over effects. We thus hypothesize the following:

H5a: *If* eigenlijk *only affects the strength of initial predictions, there will be no difference in N400 amplitude between the Eigenlijk-Coherent and Neutral-Coherent condition.*

H5b: *If comprehenders, upon encountering* eigenlijk*, revise their initial prediction and start pre-activating contrastive semantic features, there will be an increased N400 for the Eigenlijk-Coherent condition compared to the Neutral-Coherent condition.*

H6a: *If* eigenlijk *only affects the strength of initial predictions, there will be no difference in frontal positivities between the Eigenlijk-Coherent and Neutral-Coherent condition.*

H6b: *If comprehenders, upon encountering* eigenlijk, *revise their initial prediction and generate a strong, alternative lexical prediction, there will be increased frontal positivity for the Eigenlijk-Coherent condition compared to the Neutral-Coherent condition.*

In addition, combining *eigenlijk* with a discourse-coherent CW is pragmatically odd. Based on the study by Xiang and Kuperberg (2015) and EEG studies on semantic violations (Kuperberg, 2007), we argue that comprehenders could experience difficulties to integrate Coherent CWs:

H7. *If comprehenders, after encountering* eigenlijk *and a coherent CW, experience integration difficulties, there will be an increased P600 for the Eigenlijk-Coherent compared to the Neutral-Coherent condition.*

We do not expect to find a P600 effect of *eigenlijk* on Incoherent critical words.

In section 4 we will report the methods and results of the discourse completion test. We use the results and stimuli of this pretest to test the hypotheses mentioned above in an ERP experiment, which will be presented in section 5.

4. Discourse completion test

4.1 Participants

For the discourse completion task, we recruited 61 participants (16 male, 45 female) with the participant database of the Max Planck Institute for Psycholinguistics. All participants were native speakers of Dutch. Their age ranged from 18 – 30 years (M = 22.2, SD = 2.9). Participants gave full consent prior to starting the experiment and were paid 10 euros for participation. None of them had participated in any of the prior eye-tracking experiments or the corresponding pretests.

4.2 Materials and design

The stimuli were 180 short Dutch conversations in easily imaginable situations. About a quarter of the stimuli had been used before in a Visual World Paradigm study (Van Bergen, submitted), and were reused with slight adaptations. The remaining stimulus items have been newly constructed. The conversations contained colloquial language to attain an informal and natural character. Each conversation consisted of a context sentence and a question-answer pair. The last sentence, the answer, was truncated and subjects were instructed to complete it.

The context "sets the scene" and helps the reader to set up a discourse representation. The questions are suggestive and high-constraining, thereby aiming at a certain response. The most frequent question types were: (1) polar questions, (2) tag-question with 'hè' [isn't it] or 'toch' [don't you think], (3) declarative questions, mostly with 'vast' [probably] or 'zeker' [surely] and (4) negative questions. The responses contained either a neutral adverb or adverbial phrase (Neutral condition) or the discourse particle *eigenlijk* (Eigenlijk condition).[1] An example is given in Table 1, with the manipulation in boldface and the approximate English translation below.

The stimuli were counterbalanced across three lists following a Latin square design, such that each participant saw each stimulus item in only one condition (60 items per condition).[2] The order of the stimulus items within each list was randomized.

[1] The entire experimental design included a third condition with the discourse particle *inderdaad*, which falls outside the scope of this thesis rapport.
[2] This includes the Inderdaad condition.

Table 1. Example conversation.

[context]	*Diana is met haar klas van de kunstacademie een weekend naar Parijs geweest.* Diana has spent a weekend in Paris with her art academy class.
[question]	*Haar vriendin vraagt: jullie hebben zeker veel kunst gezien?* Her friend asks: you guys must have seen a lot of art?
[answer]	*Diana zegt: we zijn* ***daar / eigenlijk*** *elke dag naar een ….* Diana says: we have there / [actually] every day been to a …

4.3 Procedure

The discourse-completion task was a web-based experiment, which participants completed at home. Participants were instructed that they would read short conversations ending in an unfinished answer. They were asked to complete the answer by filling in one or several words that first came to mind. Two examples were presented before the start of the experiment.

The conversations were presented on the screen with the context, question and answer on separate lines, and a textbox below where they could type their response. There was no time limit on the task; it took participants on average 45 minutes to complete the experiment.[3]

4.4 Analysis

4.4.1 Selection of critical words

For every stimulus item, two critical words (CWs) were selected: a Target and a Competitor. The Target word is in line with the expectations of the questioner (i.e., coherent with respect to the context and the suggestive question, e.g., 'museum' in the example above), the Competitor disconfirms the expectation (i.e., it is incoherent with respect to the context and the suggestive question, i.e. 'park' in the example above). Note that we had not defined these CWs *a priori*; for every stimulus item, we selected the most frequent reply in the Neutral condition as the Target word and the most frequent reply in the Eigenlijk condition as the Competitor.[4] Three additional requirements were: (1) the Target and Competitor should be plausible given the context, (2) the contrast between the Target and Competitor should be clear and (3) every CW had to be unique, that is, it could occur only once as a CW in the entire stimuli set.

[3] Participants were able to take breaks during the experiment or even complete the experiment on another day. Therefore it is not possible to know the exact time-on-task.
[4] Usually, the most frequent word in the Neutral condition was the same as in the Inderdaad condition. However, ultimate selection was made based on the overall frequencies of responses from both the Inderdaad and Neutral sets.

4.4.2 *Coding criteria*

Responses were coded by two independent annotators. All individual responses were compared to the CWs (Target and Competitor) of the respective stimulus item and coded as a match (1) or mismatch (0). This was done at two distinct levels: at the Lemma level (is the response lexically identical or similar to the selected Target/Competitor word?) and at the Discourse level (is the response coherent or contrastive with respect to the discourse?). There were thus four coding categories: *Lemma-Target, Lemma-Competitor, Discourse-Target* and *Discourse-Competitor.*

A response was coded as a *Lemma-Target* match when it was identical to the selected Target or when there were minor lexical differences (e.g., spelling errors, abbreviations, compounds, singular-plural, verb-noun, verb tense differences etc.). Synonyms and all other words (including ambiguous responses) were coded as a mismatch. In the same way, responses have been coded as a *Lemma-Competitor* by comparing them to the selected Competitor word.

A response was coded as a *Discourse-Target* match when it was in line with the discourse and confirmed the question. A response was coded as a *Discourse-Competitor* match when it was contrastive with respect to discourse, but still plausible given the entire context. Ambiguous responses and responses that did not fit the discourse were coded as mismatches.

The inter-annotator agreement was > 95% for all categories. Disagreements were resolved through discussion.

4.5 Results

From the total of 180 stimulus items, 144 items were selected as experimental sets for the ERP study; the selection criteria can be found in Appendix A. We calculated the percentage of Target and Competitor responses (at both the Lemma and Discourse level); these cloze probabilities are presented in Table 2.

The cloze probabilities were analyzed in R with logistic mixed-effects regression analysis, using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016). The additional Inderdaad condition has been taken into account for the analysis, but only comparisons between the Neutral and Eigenlijk conditions will be reported here (see Appendix B for the complete results). The conditions have been entered into the model as fixed effects (with the Neutral condition as the intercept), and a maximal random effects structure has been used (i.e. random intercept and slopes for both subjects and stimuli, Barr, Levy, Scheepers, & Tily, 2013; see Appendix C for a specification of the models).

Table 2. Cloze probabilities resulting from the discourse completion test.

Condition	Lemma		Discourse	
	TARGET	COMPETITOR	TARGET	COMPETITOR
Neutral				
Mean [SD]	0.49 [0.50]	0.05 [0.23]	0.80 [0.40]	0.15 [0.35]
Mean range stimuli	0.05 - 1	0 - 0.40	0.38 - 1	0 - 0.45
Mean range subjects	0.27 - 0.78	0 - 0.25	0.39 - 1	0 - 0.53
Eigenlijk				
Mean [SD]	0.15 [0.36]	0.27 [0.45]	0.23 [0.42]	0.73 [0.44]
Mean range stimuli	0 - 0.80	0.05 - 0.90	0 - 0.80	0.25 - 1
Mean range subjects	0 - 0.62	0.10 - 0.46	0 - 0.71	0.37 - 0.98

The results from the analyses revealed that the cloze probabilities for the Target words were significantly lower in the Eigenlijk condition compared to the Neutral condition (Lemma: $\beta = -2.46$, SE = 0.15, $p < .001$; Discourse: $\beta = -3.51$, SE = 0.17, $p < .001$). The conversations with *eigenlijk* evoked more completions with the Competitor word; the cloze probabilities differed significantly from the Neutral condition (Lemma: $\beta = 2.47$, SE = 0.18, $p < .001$; Discourse: $\beta = 3.62$, SE = 0.18, $p < .001$).

The results from the discourse completion test thus revealed that – compared to the Neutral condition – conversations with *eigenlijk* evoked more contrastive continuations, as indicated by the reduced cloze probabilities for Target words, and the increased cloze probabilities for Competitor words. This cross-over effect is apparent at the Lemma level, and even more pronounced at the Discourse level. As such, the discourse particle *eigenlijk* thus modulates (off-line) predictions about upcoming words.

Furthermore, we would like to draw attention to several observations that follow from the results. First of all, the cloze probabilities show that the manipulation of the stimuli in terms of constraint was successful. For the conversations in the Neutral condition, ±50% of the completions contained the Target Lemma (at the discourse level this was even ±80%). The conversations in combination with *eigenlijk* were less constraining, with ±25% of the responses converging on a specific Competitor word (and ±75% at the discourse level). Thus, the stimuli can be described as medium- to high-constraining.

Secondly, as indicated by the large standard deviations and the range of stimuli and subject means, there is quite some variation – especially across items. With respect to the Neutral condition, this means that not all stimuli items point unambiguously into the direction of a specific completion. Some conversations were even completed in a contrastive manner (for one stimulus item, no less than 40% of the completions were Competitor responses). If we look at the Eigenlijk condition, we see that for some stimuli the presence of *eigenlijk* elicits hardly any Competitor responses (the lowest score being 5% at the Lemma level). This is sometimes due to the fact that participants do not converge on a specific lexical item, but provide various alternative completions (as seen by the scores at the Discourse level, which are ≥25%). To illustrate this kind of variation, consider the stimulus items (2) and (3):

(2) Q. *Hier hebben we de vorige keer goed gegeten hè?*
Our dinner over there was great last time, wasn't it?

A. *Ik herinner me* ***nog / eigenlijk*** *dat de porties daar erg …*
I still / [actually] remember that their portions were very …

(3) Q. *Je wilt een baan in de zorg gaan zoeken toch?*
You would like to look for a healthcare job, right?

A. *Ik wil* ***heel / eigenlijk*** *graag gaan werken als ….*
I would really / [actually] like to work as a ….

In (2), the particle *eigenlijk* evokes one particular alternative discourse continuation. To mark a contrast with 'big' (*groot,* the expected Target completion for this item), 90% of the participants completed the conversation in the Eigenlijk condition with the antonym 'small' (*klein*). In (3), however, to mark a contrast with the expected profession, i.e. the Target 'nurse' (*verpleegster*)*,* participants have many options. The responses in the Eigenlijk condition varied for this item (in fact, up to 14 different professions have been provided by the participants), revealing a Lemma cloze probability of 0.19, but a Discourse cloze probability of 0.81.

Additionally, the range of the means per participant indicate that some participants are more "sensitive" to the presence of *eigenlijk* than others. On the upper part of the scale, there is one participant who completed 98% of the conversations in the Eigenlijk condition with a contrastive continuation (i.e., a Competitor response at the discourse level). On the lower end, there is a participant who does so for only 37% of the conversations with *eigenlijk.* This could be explained in several ways. It could be the case that participants deliberately "ignore" *eigenlijk* (perhaps because of its high frequency in the stimuli set) and complete conversations with Target responses. It could also be due to variation across participants with respect to their social and communicative skills, as well as Theory of Mind capacities, which have been suggested to play a role in the use of *eigenlijk* (Van Bergen et al., 2011; Van Bergen, submitted).

As such, some participants might have trouble to distill the interpersonal meaning from *eigenlijk*, and/or to subsequently come up with an alternative, but plausible continuation.[5]

[5] To capture individual variation in social and communicative skills in the ERP study, we asked subjects to fill out the *social skill* and *communication* subscales of the Autism-Spectrum Quotient questionnaire (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). These findings, and the possible correlation with the ERP results, fall outside the scope of this thesis rapport.

5. ERP study

5.1 Participants

Forty participants (13 male, 27 female) were recruited with the participant database of the Max Planck Institute for Psycholinguistics. All participants were native speakers of Dutch and right-handed. Their age ranged from 18 – 28 years (M = 22.2, SD = 2.1). Participants gave full consent prior to starting the experiment and were paid 18 euros for participation. None of them had participated in any of the prior eye-tracking experiments or discourse completion tests. Three subjects were excluded from data analysis: two because of extensive alpha waves and one because of problems with referencing. We report data from the remaining 37 participants.

5.2 Materials and design

The materials for the reading task consisted of the 144 conversations which had been selected from the discourse completion test. As described in section 4, we pretested the items in two conditions (the response contained a neutral adverb or the particle *eigenlijk*). From the participant's responses we selected two different critical words (CWs) for every item, one which is discourse-coherent (the Target) and one which is discourse-incoherent (the Competitor). As such, we attained a 2 (Particle: Eigenlijk or Neutral) by 2 (Coherence: Coherent or Incoherent) design. The Coherent and Incoherent CWs were comparable in terms of length (mean 7.0 and 6.7 letters, respectively) and frequency (both mean 1.2 logarithmic frequency; CELEX corpus). An overview with an example item is presented in Table 3. This table also lists the cloze probabilities.

The entire research design contained two additional conditions, for which an extra 72 items had been constructed and pretested. These conditions fall outside the scope of this thesis rapport and will not be part of the reported results.[6] The complete stimulus set for this experiment thus contained 216 short conversations. The items were counterbalanced across four lists following a Latin square design, such that each participant read each conversation in only one condition (36 items per condition). The order of items within a list was semi-randomized (such that the same condition appeared maximally three times in a row).

[6] These additional experimental sets were weak-constraining (the most common completion for a certain item in the discourse completion task always had a probability < 0.2). There were two conditions: a neutral condition (with a neutral adverb/adverbial phrase; Neutral-Weak) and a condition with the particle *eigenlijk* (Eigenlijk-Weak). The CW in these conditions was always unexpected (cloze probability < 0.05).

Table 3. Example stimuli and cloze probabilities.

Condition	Example	Lemma	Discourse
	Diana has spent a weekend in Paris with her art academy class. *Her friend asks: you guys must have seen a lot of art?* *Diana says:*		
Neutral-Coherent	We zijn daar elke dag naar een **museum** geweest. *We have there every day been to a museum.*	0.49 [0.50]	0.80 [0.40]
Neutral-Incoherent	We zijn daar elke dag naar een **park** geweest. *We have there every day been to a park.*	0.05 [0.23]	0.15 [0.35]
Eigenlijk-Coherent	We zijn eigenlijk elke dag naar een **museum** geweest. *We have [actually] every day been to a museum.*	0.15 [0.36]	0.23 [0.42]
Eigenlijk-Incoherent	We zijn eigenlijk elke dag naar een **park** geweest. *We have [actually] every day been to a park.*	0.27 [0.45]	0.73 [0.44]

Note: Means are shown with standard deviation in square brackets. The critical manipulations in the examples sentences are underlined (Particle) or bold (CW). Cloze probabilities are presented as the proportion of total responses from 61 participants (see section 4).

5.3 Procedure

The experiment was conducted at the Max Planck Institute for Psycholinguistics in Nijmegen. After EEG and GSR preparation[7], participants were tested in a dimly lit soundproof booth. They were seated in a comfortable chair in front of a computer screen at a viewing distance of approximately 100 cm. Participants were instructed that they would be reading short conversations on the screen, which consisted of a context sentence, question and answer. They were told that the context and question would be presented in full and the answer word-by-word. They could read the context and question at their own pace, pressing a button to move on to the next screen. Participants were instructed to sit still and avoid blinking during the word-by-word presented answer. They were informed that their understanding of the conversations would be tested at the end of the reading task, so they would remain alert throughout the task.

The stimuli were presented with a Presentation script (Neurobehavioral systems™), in a black font (Lucida console, 26-point size) and centered on a white background. Each answer was preceded by a fixation cross (1000 ms). The first part of the answer (e.g. "Jan says") remained on the screen for 800 ms. The subsequent presentation time of the words in the answer sentence was variable to attain natural reading times (cf. Nieuwland & Van Berkum, 2006). Word duration was computed as (number of letters * 30) + 190 ms, with a maximum of 400 ms. However, we made two exceptions to avoid spurious ERP effects due to differences in

[7] During the ERP experiment, we also collected Galvanic Skin Response data to measure arousal. The results will be analyzed later and fall outside the scope of this thesis report.

presentation times: the word in the Particle area[8] (neutral adverb or adverbial phrase versus *eigenlijk*) and the CW had a non-variable duration of 400 ms. The final word of the sentence appeared with a period and was presented for 800 ms, after which the next trial would start automatically. The inter-word-intervals was 150 ms.

Participants started with a practice block of four trials, after which they could ask clarification questions and/or received feedback before starting the experiment. The total of 216 trials was divided into 6 blocks separated by self-timed breaks. The practice and the experimental blocks together lasted about 45 – 60 minutes (depending on the reading pace of the participant).

5.4 Apparatus

The EEG response was recorded from 31 cap-mounted Ag/AgC1 electrodes (actiCAP, Brain Products GmbH; see Figure 1 for montage).[9] Five electrodes were placed on the midline sites Fz, FCz, Cz, Pz and Oz. Twenty-six electrodes were placed over the lateral sites Fp1/2, F3/4, F7/8, FC1/2, FC5/6, C3/4, T7/8, CP1/2, CP5/6, P3/4, P7/8, O1/2 and PO9/10. Two separate electrodes were placed on the left and right mastoids (outside of the cap). During EEG recording, all electrodes were referenced to the left mastoid. Blinks were monitored by two additional electrodes placed above and below the left eye. Horizontal eye movements were monitored through two additional electrodes placed at the outer canthus of both eyes. The signal was amplified by BrainAmp DC amplifiers, filtered online with a band-pass filter between 0.1 and 1000 Hz and digitized at 500 Hz.

[8] Some of the items contained two or three words in the Particle area rather than one word (e.g. 'heel graag', 'de hele week'). In those cases, the first word(s) had a variable presentation time, and the last word was presented with the fixed 400 ms presentation time.

[9] We have changed the EEG montage set-up after data had already been collected from 17 participants. The set-up used for participant 18 onwards is reported in this section. In the initial set-up we used 27 cap-mounted electrodes instead of 31; the four electrodes for scalp sites Fp1/2 and PO9/10 were used to measure eye-movements.

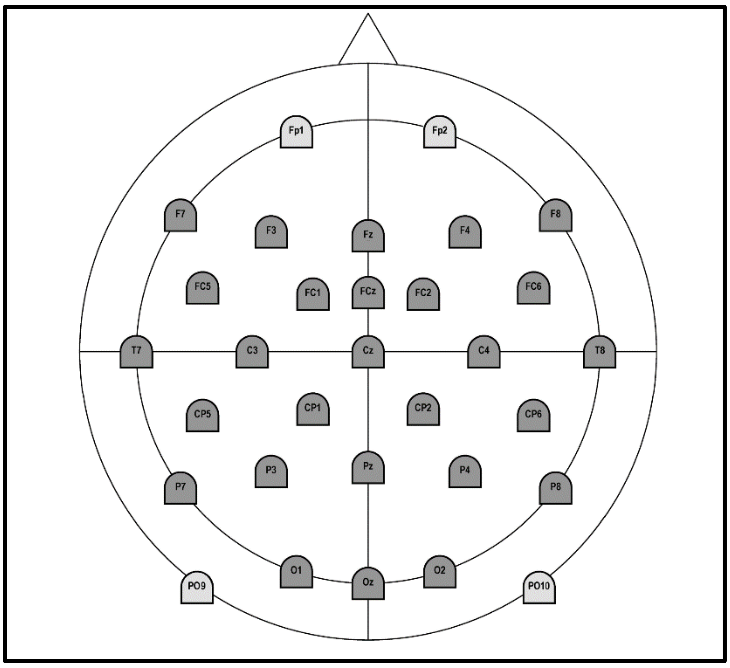

Figure 1. Electrode montage for the EEG experiment. Electrodes Fp1/2 and PO9/10 have been used for a subset of the participants (see footnote 8).

5.5 Data pre-processing

The software package Brain Vision Analyzer 2TM (Brain Products GmbH) was used to analyze the waveforms. First, the EEG signal was re-referenced to the mean of the left and right mastoids. EOG signals were re-referenced as well: vertically to the mean of the electrodes below and above the left eye, horizontally to the mean of the electrodes on the outer canthi. Then, EEG activity was filtered with zero phase shift Butterworth IIR filters (low cutoff: 0.1 Hz, 2nd order; high cutoff: 20 Hz, 4th order). Similarly, EOG activity was filtered (high cutoff: 20 Hz, 4th order). Segments were extracted from -200 ms until 1000 ms relative to CW onset and baselined to a 200 ms pre-onset baseline. We proceeded with semi-automatic artifact rejection based on the following criteria. Segments with vertical eye artifacts (> ± 50 µV) or horizontal eye artifacts (>± 40 µV) were screened. Remaining segments with artifacts on the EEG signals were screened when it exceeded the limits of -100 and 100 µV or when it contained voltage steps higher than 50 µV/ms. Segments were ultimately discarded upon visual inspection, with no asymmetry over conditions. The overall segment loss was 9%; for all four conditions an average of 33 trials (range: 25 – 36 trials) per participant remained.

5.6 Analysis

The segments that remained after preprocessing were averaged per participant and per condition. To capture the ERP effects that we were interested in (the N400 and the frontal and posterior late positivities), we chose two time windows. For the N400, we used the canonical 300-500 ms time window. The later positivity effects (in particular the frontal positivity effect) are less well established, but both are most typically measured 600-900 ms after word onset (cf. the systematic survey by Petten & Luka, 2012). By choosing this time window we also avoided component overlap with the earlier N400 effect.

The mean amplitude values (averaged over items per condition) in these latency ranges were submitted to repeated measures analyses of variance, using the Greenhouse-Geisser correction for univariate F tests with more than one degree of freedom (we report the original df). The N400 time window was first inspected in an overall analysis with all 27 electrodes,[10] thus defining a 2 (Coherence: Coherent, Incoherent) x 2 (Particle: Neutral, Eigenlijk) x 27 (Electrode) design. We defined two separate (non-overlapping) regions of interest (ROI) for the positivity effects, based on the frontally and posteriorly distributed effects that have been previously found in the literature (e.g., Federmeier et al., 2007; Kim & Osterhout, 2005; Thornhill & Petten, 2012): one frontal ROI (FC1/2, FC5/6, F3/4, F7/8, Fz, FCz) and one posterior ROI (CP1/2, CP6/6, P3/4, P7/8, O1/2, Pz, OZ). For both regions we conducted an overall analysis with all 10 or 12 electrodes, thus defining 2 (Coherence: Coherent, Incoherent) x 2 (Particle: Neutral, Eigenlijk) x 10 or 12 (Electrode) designs.

Subsequently, to explore any interaction effects, the topography was explored in a mean quadrant analysis involving the left anterior electrodes (F3, F7, FC1, FC5), the right anterior electrodes (F4, F8, FC2, FC6), the left posterior electrodes (CP1, CP5, P3, P7, O1) and/or the right posterior electrodes (CP2, CP6, P4, P8, O2), with Hemisphere (Left, Right) and/or Anteriority (Anterior, Posterior) as factors.

Since we had two additional frontal electrodes for the second half of the participants (N = 21), we carried out the analysis for the frontal ROI again for this subset with electrodes Fp1/2 included.

[10] These are the 27 electrodes for which we have data points for all participants; the sites Fp1/2 and PO9/10 are therefore not part of this overall analysis.

5.7 *Results*

5.7.1 N400: 300-500 ms

The grand-average waveforms elicited by the coherent ('museum') and incoherent ('park') critical words are displayed in Figure 2 for the Neutral (A) and Eigenlijk (B) conditions separately (the grand-average plots with all conditions are added as Appendix D). As expected, incoherent words elicited negativities relative to coherent words, lasting from approximately 300 to 500 ms, as is typical for N400 effects. We will report main effects and interactions only when they involve the factors Coherence and/or Particle. Nonsignificant effects are reported only when the lack of significance is relevant for the hypotheses of the present study.

The overall analysis of variance over the whole scalp (see Table 4) revealed an effect of Coherence, indicating that Incoherent critical words (0.63 μV) elicited more negative mean amplitudes than Coherent critical words (1.75 μV), as visualized in Figure 3, panel A. This effect of Coherence varied in size across the 27 electrodes (Coherence x Electrode interaction). There is no main effect of Particle. The interaction of Coherence x Particle does not reach significance, nor does the interaction of Particle x Electrode or the three-way interaction Prediction x Particle x Electrode.

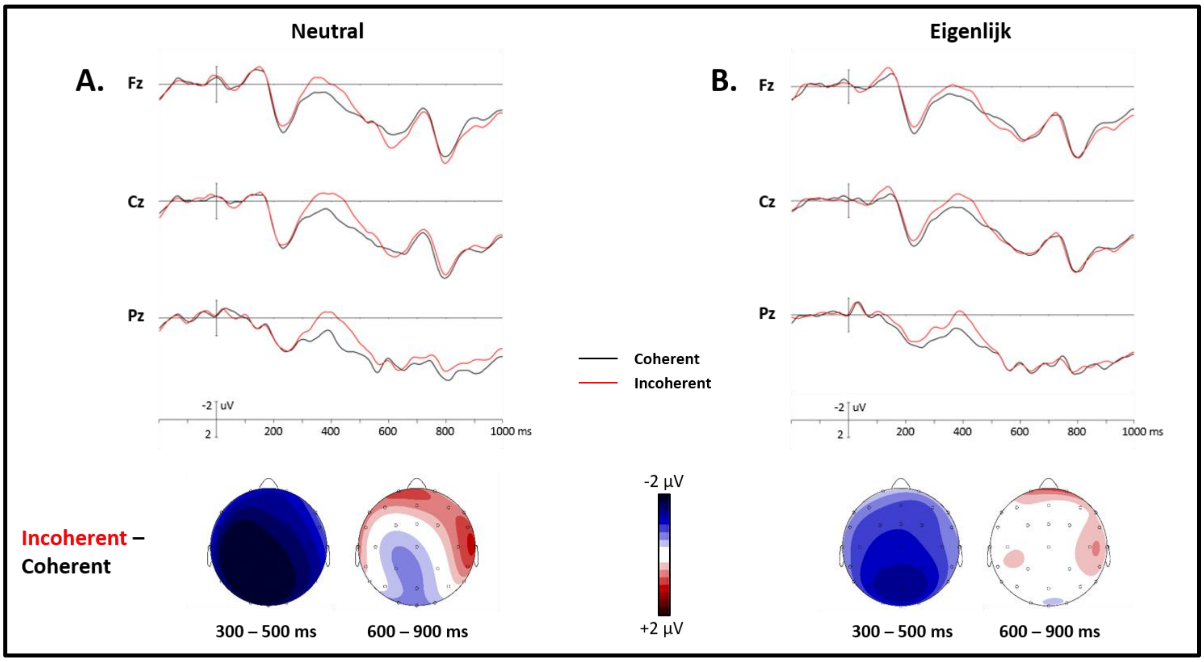

Figure 2. Grand-averaged (N = 27) waveforms (negative plotted up) to critical words (CWs) showing effects of electrodes Fz, Cz and PZ.
Panel A: waveforms to CWs in Neutral-Coherent (black) and Neutral-Incoherent (red) conditions.
Panel B: waveforms to CWs in Eigenlijk-Coherent (black) and Eigenlijk-Incoherent (red) conditions.
Scalp topographies show differences in ERPs between Incoherent and Coherent critical words (Incoherent minus Coherent) between 300–500 ms (N400) and 600–900 ms (late positivities).

As can be seen from the scalp topographies in Figure 2, the effect of Coherence seems to be distributed over the whole scalp, with a maximum over centro-parietal electrodes. To further investigate the distribution of this effect, we conducted a quadrant analysis (see the statistics in Table 4). The results again reveal a main effect of Coherence. Furthermore, the Coherence x Anteriority interaction is trend-wise significant, suggesting that the difference between Coherent and Incoherent critical words was slightly larger over posterior (1.23 μV difference) than anterior regions (0.89 μV difference), which is in line with previously reported N400 effects. Follow-up analyses reveal that the effect of Coherence is significant in both anterior ($F(1, 36) = 21.47$, $p < .001$) and posterior quadrants ($F(1, 36) = 30.93$, $p < .001$).

Table 4. Omnibus ANOVAs for the N400 time window (300-500 ms)

Overall analysis (27 electrodes)			
Effect	df	*F*	p
Coherence	(1, 36)	34.74	<.001
Particle	(1, 36)	0.96	.335
Electrode	(26, 936)	8.90	<.001
Coherence × Particle	(1, 36)	2.04	.162
Coherence × Electrode	(26, 936)	4.71	<.001
Particle × Electrode	(26, 936)	1.04	.396
Coherence × Particle × Electrode	(26, 936)	0.46	.762
Mean quadrant analysis (4 quadrants)			
Effect	df	*F*	p
Coherence	(1, 36)	32.60	< .001
Anteriority	(1, 36)	11.97	.001
Hemisphere	(1, 36)	14.89	<. 001
Coherence × Anteriority	(1, 36)	3.39	0.074
Coherence × Hemisphere	(1, 36)	1.64	0.209
Anteriority × Hemisphere	(1, 36)	3.33	0.076
Coherence × Anteriority × Hemisphere	(1, 36)	< .01	0.957

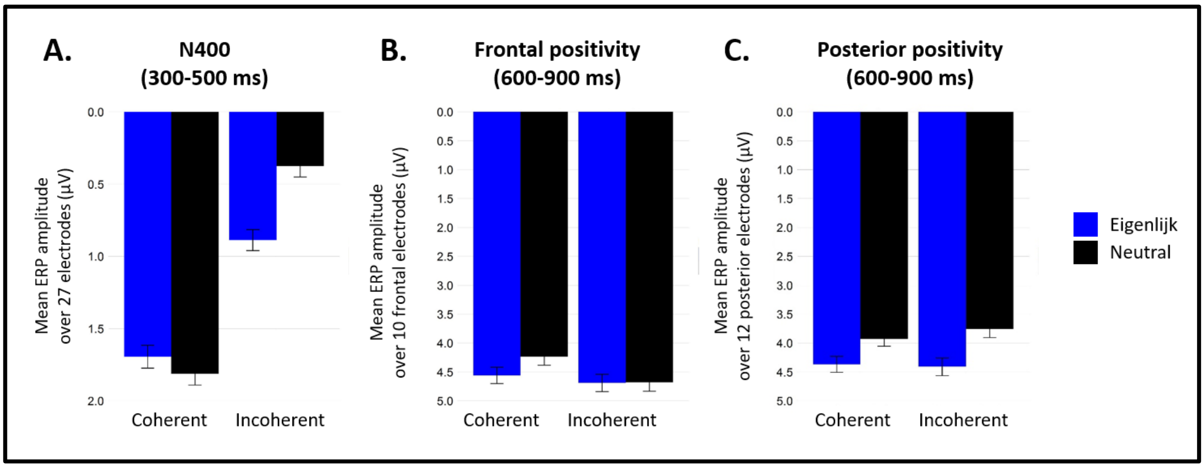

Figure 3. Bar plots conveying the ERP mean amplitudes (N=37) per region of interest included in each analysis in the respective time windows. Error bars indicate standard error. Note that panel B conveys the mean amplitudes over 10 frontal electrodes (this excludes Fp1/2).

5.7.2 Frontal positivity: 600-900 ms

For the second time window (600-900 ms), we first inspected frontal positivity effects by focusing on the anterior region. The overall analysis of variance over the anterior region (with 10 electrodes, see Table 5 for the statistics) did not reveal a main effect of Coherence nor a main effect of Particle. None of the interactions were significant either.

Note, however, that this analysis has been conducted with four electrodes per quadrant, as all 27 participants had data points for these electrodes. However, it does appear from Figure 2 that there is some sort of late positivity effect in the anterior region, which seems to be maximal in the most frontal electrodes. Therefore, we conducted an additional analysis with the subset of the participants for which data was collected at sites Fp1 and Fp2. In this overall analysis (N=21, see Table 5 for statistics) over the anterior region, the main effect of Coherence is not significant, but there does seem to be a trend (p = .086).

Table 5. Omnibus ANOVAs for the frontal late positivity (600 – 900 ms)

Overall analysis (10 electrodes)			
Effect	df	*F*	p
Coherence	(1, 36)	1.78	.190
Particle	(1, 36)	0.52	.477
Electrode	(9, 324)	66.66	<.001
Coherence × Particle	(1, 36)	0.39	.536
Coherence × Electrode	(9, 324)	1.09	.365
Particle × Electrode	(9, 324)	0.52	.737
Coherence × Particle × Electrode	(9, 324)	0.78	.541
Overall analysis (12 electrodes, including Fp1/2)*			
Effect			
Coherence	(1, 20)	3.27	.086
Particle	(1, 20)	<.001	.995
Electrode	(11, 220)	28.03	<.001
Coherence × Particle	(1, 20)	0.08	.784
Coherence × Electrode	(11, 220)	0.92	.469
Particle × Electrode	(11, 220)	0.47	.810
Coherence × Particle × Electrode	(11, 220)	0.53	.699

* Note: this analysis has been conducted with a subset of the participants (N=21).

5.7.3 Posterior positivity: 600-900 ms

As can be seen from the waveforms Figure 3, the Eigenlijk conditions elicited more positive amplitudes compared to the Neutral conditions in the later time-window (600-900 ms). Based on the scalp topographies, this effect seems to be maximum over posterior electrodes, and right-lateralized in the Eigenlijk-Coherent condition.

The overall analysis of variance (see Table 6) over the posterior region indeed revealed a main effect of Particle, indicating that critical words preceded by *eigenlijk* elicited a more positive mean amplitude (4.39 uV) compared to critical words in the Neutral condition (3.85 uV; see Figure 3, panel C). This effect does not seem to vary across the 12 electrodes, as the Particle x Electrode interaction does not reach significance. The ANOVA did not reveal a significant main effect of Coherence, nor an interaction of Coherence × Particle or Coherence × Particle × Electrode. However, the interaction of Coherence × Electrode is significant.

To further explore the distribution of the significant effects, a mean quadrant analysis of variance was conducted for the two posterior quadrants (see Table 6 for the statistics). Again, there is a main effect of Particle. The main effect of Coherence is not significant, and neither are any of the interaction effects with this factor.

However, the interaction of Particle × Hemisphere approaches significance, suggesting that the effect of Particle is slightly larger in the right posterior quadrant (0.64 uV difference between Eigenlijk and Neutral conditions) compared to the left posterior quadrant (0.36 uV difference). Follow-up analyses reveal that the effect of Particle is significant over the right posterior electrodes ($F(1, 36) = 10.99$, p = .002), but not in the left posterior region ($F(1, 36) = 3.35$, p = . 075).

Table 6. Omnibus ANOVAs for the posterior late positivity (600-900 ms)

Overall analysis (12 posterior electrodes)			
Effect	df	*F*	p
Coherence	(1, 36)	0.09	.762
Particle	(1, 36)	8.95	.005
Electrode	(11, 396)	71.81	<.001
Coherence × Particle	(1, 36)	0.17	.685
Coherence × Electrode	(11, 396)	5.17	<.001
Particle × Electrode	(11, 396)	1.69	.136
Coherence × Particle × Electrode	(11, 396)	1.50	.197
Mean quadrant analysis (2 posterior quadrants)			
Effect	df	*F*	p
Coherence	(1, 36)	< .01	0.962
Particle	(1, 36)	7.58	0.009
Hemisphere	(1, 36)	4.10	0.050
Coherence × Particle	(1, 36)	0.09	0.766
Coherence × Hemisphere	(1, 36)	1.72	0.198
Particle × Hemisphere	(1, 36)	4.06	0.052
Coherence × Particle × Hemisphere	(1, 36)	1.89	0.178

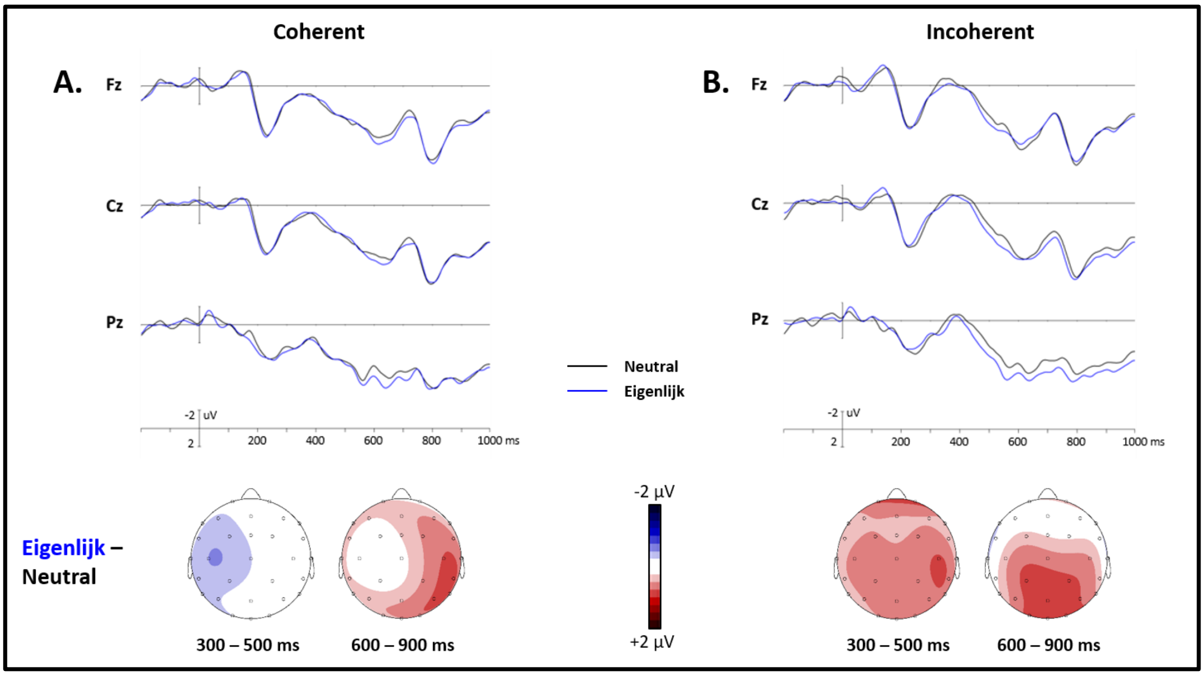

Figure 3. Grand-averaged (N = 27) waveforms (negative plotted up) to critical words (CWs) showing effects of electrodes Fz, Cz and PZ.
Panel A: waveforms to CWs in Neutral-Coherent (black) and Eigenlijk-Coherent (blue) conditions.
Panel B: waveforms to CWs in Neutral-Incoherent (black) and Eigenlijk-Incoherent (blue) conditions.
Scalp topographies show differences in ERPs between the two Particle conditions (Eigenlijk minus Neutral) between 300–500 ms (N400) and 600–900 ms (late positivities).

6. Discussion

6.1 N400: 300-500 ms

Focusing first on the N400 time window, we found a main effect of Coherence, which is in line with our hypothesis and previous studies showing that critical words which are unexpected given a wider discourse elicit N400 effects (e.g., the work by Van Berkum and colleagues; see Van Berkum, 2009 for a review). Furthermore, the effect found in the present study is similar to what has been reported in previous studies, both in terms of time-course (peaking at ± 400 ms after critical word onset) and distribution (significant over the whole scalp, but maximum over posterior electrodes) – though there is no sign of a (slight) right hemisphere bias (e.g. Kutas & Federmeier, 2011; Van Berkum, Hagoort, & Brown, 1999). However, rather than merely replicating previously reported N400 effects, the current study adds to the field as the effect has been found in a conversational setting. Of course, reading experimentally controlled sentences is far from natural dialogue, yet it does go beyond the discourse level, as there are two speakers involved – each with their own discourse representation. The present study thus shows that the ease of semantic processing of information in a certain response depends on how well it fits within that particular dialogue, given the context information and the expectations of the question-asker (as inferred from the suggestive question). As such, we contributed to the pioneering work on dialogue using ERPs, and affirm the conclusion by Bögels et al. (2015) that "the N400 can not only index expectations within an utterance or a monologic discourse but can also reflect expectations over the turns of different speakers in a conversation" (p. 12).

However, counter to our initial expectations, the N400 effect did not appear to be modulated by the presence of *eigenlijk*. We argued that the particle *eigenlijk* can function as a cue for upcoming unexpectedness, and we thus expected to find an attenuated N400 effect for the Eigenlijk-Incoherent condition relative to the Neutral-Incoherent condition. This hypothesis was based on (a) the findings from the Visual World Paradigm study by Van Bergen (submitted), which revealed that listeners immediately integrate the information encoded in *eigenlijk* to modulate anticipatory looks to possible referents and (b) several ERP studies which showed that specific linguistic expressions can facilitate the processing of unexpected information, as indicated by the N400 component (e.g. *even so,* Xiang & Kuperberg, 2015). We have listed these two reasons, since it does not necessarily follow from the VWP study alone that *eigenlijk* will affect predictions in a similar manner in the present ERP study. In the VWP study, listeners immediately shifted their attention to the alternative referent (the Competitor) upon encountering *eigenlijk.* However, by setting up a visual world with only four pictures, expectations about upcoming referents were highly restricted. In the present study, this was not the case. Consider again the question-answer pair in (2) below, which was a stimulus item in the present ERP study:

(3) Q. *Je wilt een baan in de zorg gaan zoeken toch?*
You would like to look for a healthcare job, right?

A. *Ik wil **heel / eigenlijk** graag gaan werken als ….*
I would really / [actually] like to work as a ….

If we assume that comprehenders were affected by the presence of *eigenlijk*, and thus expect a contrastive completion to follow (i.e., not 'nurse', but another profession), there is a whole array of possibilities. That is, the cloze probability at the Lemma level is rather low (in this case 0.19; the mean for the Eigenlijk-Incoherent condition is 0.27). We know from previous studies that the N400 effect varies as a function of cloze probability (e.g., Kutas & Hillyard, 1984), and therefore the semantic processing of such weakly expected incoherent critical words might not be facilitated. However, the cloze probability of the Eigenlijk-Incoherent condition is still higher than in the Neutral-Incoherent condition (0.27 versus 0.05), and we thus expected to see this in a modulation of the N400 effect. Based on visual inspection of Figures 2 and 3, there does seem to be a trend in this direction (i.e., less negative amplitudes for the Eigenlijk-Incoherent compared to Neural-Incoherent condition), but the analysis did not reveal a significant interaction of Coherence by Particle.

One possible explanation for the absence of a significant interaction effect is that comprehenders were influenced by the presence of the Eigenlijk-Coherent condition, in which *eigenlijk* is pragmatically odd. One could speculate that the listeners refrained from updating their expectations upon encountering *eigenlijk*, as they have learned that in 50% of the cases a coherent critical word will still follow. This possibility is intriguing and could be explored in future research. However, it should be noted that even though *eigenlijk* apparently does not facilitate the semantic processing of unexpected information, this does not mean that upon encountering *eigenlijk* comprehenders have stopped anticipating upcoming information all together.[11]

For the coherent critical words, even though the Eigenlijk-Coherent responses were pragmatically odd, they did not elicit more negative amplitudes (compared to the Neutral-Coherent condition). There are two possible interpretations for this finding. First of all, as

[11] When comprehenders do not (or hardly) anticipate upcoming information – as is the case for the weak-constraining conditions that we included in the design – there is a much more negative N400 waveform (see Appendix D for a plot with the waveforms of all six conditions).

mentioned before, it could be that the presence of both Eigenlijk-Coherent and Eigenlijk-Incoherent conditions made the participants refrain from modulating their expectations, and thus "sticking to" their prediction for the coherent critical word.

However, the presence of *eigenlijk* before a coherent critical word does have an effect in the later time window (600-900 ms), thus it does not seem to be the case that comprehenders completely disregard the information that this discourse particle conveyed. Instead, we believe that the absence of the N400 can be accounted for if we take this component to reflect memory retrieval processes (cf. Brouwer, Fitz, & Hoeks, 2012). Retrieving the lexical information associated with a coherent critical word (e.g., 'nurse') from long-term memory could be facilitated by the activation of semantic features of the preceding context (e.g., 'healthcare', 'job') – and it is likely that this kind of automatic priming is independent of the presence of the discourse particle *eigenlijk.* It could well be that *eigenlijk* asserts its influence in another manner which still requires some kind of prolonged processing, as indicated by the later positivity effect. Relevant to note is that this pattern – a P600 effect in absence of an N400 effect – has been reported in other ERP studies as well, for example with respect to semantic violations (Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kolk, Chwilla, Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003), inferential processing (Burkhardt, 2007) and irony processing (Regel et al., 2011; Spotorno et al., 2013). This will be further discussed in section 6.3.

6.2 Frontal positivity: 600-900 ms

In the later time window (600-900 ms), we had separate hypotheses for anterior and posterior regions. For the anterior region, we expected to find a positivity effect for discourse incoherent CWs which violate a strong prediction (thus reflecting a prediction error) – an effect which could potentially be attenuated when preceded by *eigenlijk*. The results did not reveal such an interaction effect, though in the analysis which included electrodes Fp1 and Fp2 (with a subset of the participants) the main effect of Coherence was trend-wise significant. However, we should be careful to over-interpret these results given the low statistical power with such a small sample size (N=21).

We furthermore believe the effect could have been too small to detect due to the nature of the stimuli items. The mean cloze probability (at the Lemma level) for the Neutral-Coherent condition in our stimuli set was rather low (0.49) compared to the other studies in which the effect has been reported (0.90 in DeLong et al., 2014; 0.78 in Thornhill & Petten, 2012). Even though our contexts and questions were high-constraining when measured on the Discourse level (0.80 on average for the Neutral-Coherent condition), this does not affect the frontal positivity component, as it has been found to reflect the disconfirmation of specific *lexical* predictions (Thornhill & Van Petten, 2012).

The present study does not provide any evidence that the particle *eigenlijk* influences late frontal positivities. Initially, we put forward two possible ways in which it could asserts its effect. First of all, we hypothesized that *eigenlijk* could function as a warning for upcoming unexpected information, and thus prevent the occurrence of a prediction error. However, as just mentioned, the effect of Coherence by itself might have been very small as not all contexts were highly lexically constraining, which makes it hard to detect an interaction with Particle – even more so given the variable cloze probabilities at the Lemma level in the Eigenlijk-Incoherent condition (as discussed in more depth in section 4.5). Therefore, we would not want to exclude the possibility that *eigenlijk* could assert an influence on the frontal positivity component in the Eigenlijk-Incoherent condition, but this clearly warrants further research.

The second possibility, which was more of an open question, would be to find a cross-over effect of *eigenlijk*. That is, that *eigenlijk* could evoke new, contrastive predictions, which would result in a prediction error in the Eigenlijk-Coherent condition. The present results contradict this (though we should again be careful given the small sample size for which we have all data from all frontal electrodes). Nonetheless, some positivity is visible in the frontal region for the Eigenlijk-Coherent condition (see panel B of Figure 2 and panel A of Figure 3). We believe, however, that this could be part of a more widely distributed pragmatic P600 (cf. Spotorno et al., 2013). We will further elaborate on this in the next section on the posterior positivity.

6.3 Posterior positivity: 600-900 ms

For the posterior region, we hypothesized that if *eigenlijk* modulates predictions about upcoming information in an incremental manner, then encountering the initially expected critical word would result in (semantic) integration difficulties, as reflected by a P600 effect (cf. Xiang & Kuperberg, 2015). At first sight the results appear to support this reasoning, as we found increased P600 amplitudes for coherent words preceded by *eigenlijk* compared to words preceded by a neutral adverb. However, there are two objections to this interpretation. First of all, the N400 and frontal positivities results did not provide evidence that *eigenlijk* directly modulates predictions. Secondly, the presence of *eigenlijk* does not only evoke increased posterior positivity for coherent critical words, but also for incoherent critical words. If the P600 reflects difficulties to integrate an incoming word, it is not clear why the incoherent critical word preceded by *eigenlijk* would elicit more processing effort compared to the neutral condition. In both cases the initial discourse representation (e.g., someone going to Paris with her art class, where she probably visited several musea) needs to be updated with the unexpected information (e.g., she visited parks).

However, an alternative interpretation for the main effect of *eigenlijk* is not straightforward either, as there is a lack of consensus in the current literature on the underlying functionality of late positive effects. As a general starting point we will take the integration view on the P600 by Brouwer, Fitz and Hoeks (2012), as previously outlined in section 2.2.2. According to this account, a P600 effect reflects the additional effort to integrate certain information into a discourse representation. Note that this does not necessarily concern updating in terms of propositional content (which can be computed from the literal meaning of the utterance), but that it could also be information that the comprehender obtains after pragmatic inferencing – the so-called *speaker meaning* or *intended meaning*. Accordingly, Brouwer et al.'s account predicts an increase in P600 amplitude "whenever recovering what the speaker or reader means requires intensive pragmatic processing" (p. 139).

With regards to this notion of speaker meaning, we believe that speakers do not only want to exchange information (e.g., that he or she visited a lot of parks in Paris), but they might also want to convey that the expectations of the addressee were legitimate given his or her discourse representation, and thus signal their attitude and social intentions. This is wat the speaker does by using *eigenlijk*, and it might be the reflection on this on the part of the comprehender which is reflected in the P600 effect that we found. In other words, the posterior positivity effect could be related to some kind of reflection on the socio-pragmatic function of *eigenlijk*, as the comprehender tries to detect the speaker's attitude and intention. Our results indicate that such intensive processing comes about irrespective of the coherence of the critical word, and we thus believe it could reflect processing in the domain of social cognition, rather than the integration of the propositional content of the utterance.

This additional processing effort could be similar to the pragmatic inferencing related to irony (Regel et al., 2011; Spotorno et al., 2013) and indirect requests (Coulson & Lovett, 2010), for which similar late positivities have been reported. For irony it has been argued previously that it involves detecting the speaker's *attitude* and *intentions*, and thus requires social cognitive resources or Theory of Mind (Spotorno et al., 2013; Nicola Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012; Wilson, 2009). Interestingly, apart from a P600 effect for irony processing, Spotorno et al. (2013) also conducted a time frequency analysis (TFA) and found increased power in the gamma band. They propose that this reflects "the engagement of social cognitive processes" (p. 8). Furthermore, interpreting the main effect of *eigenlijk* as reflecting pragmatic inferencing also makes sense with regard to its non-canonical distribution; instead of being largest over posterior electrodes with bilateral symmetry, it appeared to have a right hemisphere bias. Previous studies have also reported right-biased late positivities, for example in relation to jokes (Coulson & Lovett, 2004; Coulson & Kutas, 2001) and indirect requests (Coulson & Lovett, 2010).

To sum up, we have proposed that the P600 effect for critical words which are preceded by *eigenlijk* (compared to neutral adverbs) reflects pragmatic inferencing to uncover what the speaker intended to convey with the utterance, and that such additional cognitive processes occur irrespective of the coherence of the critical word. This might mean that the use of *eigenlijk* does not only require Theory of Mind efforts on the part of the speaker, but also on the part of the comprehender as he or she tries to establish the social significance of the utterance in which it was used. We could speculate about the exact nature of these pragmatic inferences, and argue that, subsequently, the coherent and incoherent critical words might be integrated into the discourse representation in a different way. For example, upon encountering an incoherent word, the comprehender might acknowledge or even appreciate the face-saving act of the speaker, whereas the coherent word might evoke some confusion ("why did the speaker use *eigenlijk*?"). Unfortunately we cannot draw any conclusions related to this issue based on the present study, but clearly the results give rise to many questions in need of further investigation. For an optimal comparison between the underlying mechanisms of the P600 for different uses of *eigenlijk*, as well for comparing diverse pragmatic phenomena, TFA would be an interesting direction for further research. To establish the extent to which Theory of Mind processes are involved, further research could be undertaken with participants who vary in their mind-reading skills, for example by comparing Autism Spectrum Disorder patients with healthy controls.

Finally, it is important to keep in mind that the subjects in our ERP reading task were passive "overhearers" of the conversation, rather than the person to whom the *eigenlijk* was addressed. We would like to argue that the current results suggest that overhearers are also sensitive to the intersubjective information conveyed by *eigenlijk*. This is in line with fMRI studies which have shown that, while listening to indirect speech, "mentalizing regions" (associated with Theory of Mind processing) showed increased activity in the brains of both addressees and overhearers (Bašnáková, Van Berkum, Weber, & Hagoort, 2015; Bašnáková, Weber, Petersson, Van Berkum, & Hagoort, 2014). Nonetheless, the actual addressees could take the use of *eigenlijk* and its social significance more personal, which might be reflected differently in neural mechanisms and brain potentials. This could be explored with more interactive settings in future research.

7. Conclusion

The present study set out to investigate how the expectation-managing discourse particle *eigenlijk* affects on-line language comprehension in a conversational context. More specifically, we were interested in the effect of *eigenlijk* on predictive processing, by examining its influence on the pre-activation and integration of information which is coherent or incoherent given the discourse content. To this end, we set up a large stimulus set consisting of written mini-conversations, which we used to conduct an off-line discourse completion test and an on-line ERP reading experiment.

We found that *eigenlijk* modulates off-line sentence completions; it evoked more contrastive continuations of the conversation compared to the neutral condition. As has been shown by a recent eye-tracking study (Van Bergen, submitted), *eigenlijk* can also modulate expectations about upcoming information during incremental language comprehension. To study the influence of *eigenlijk* on the pre-activation and integration of information during sentence processing, we have turned to ERPs. As a starting point we used the well-established finding that words which are incoherent given a certain discourse evoke N400 effects, as well as frontal positivity effects in the case of disconfirmation of specific lexical predictions. We hypothesized that if *eigenlijk* is immediately integrated to modulate the pre-activation of upcoming information, that this would result in an attenuated N400 for incoherent critical words. We also argued that it could induce comprehenders to "let go of" any initial specific prediction, and might thus prevent prediction error, as reflected by an attenuated frontal positivity effect.

Our results for the N400 confirmed findings from previous studies: the semantic processing of coherent critical words was facilitated compared to incoherent critical words. However, we found no significant differences between conversations with *eigenlijk* compared to neutral ones, and therefore argued that the semantic memory retrieval of incoherent words was not facilitated. Comprehenders thus do not seem to be affected by *eigenlijk* when it comes to the pre-activation of upcoming information; they do not take it as a cue to rapidly come up with new predictions (or they might not have been able to do so with enough certainty). Furthermore, we found no significant frontal positivity effects for the incoherent critical words, which could be due to the fact that discourse contexts were not constraining enough and/or because of a lack of statistical power – whereupon no interaction effects with *eigenlijk* were found either.

We also argued that the presence of *eigenlijk* in combination with the initially expected and discourse-coherent information is in a sense pragmatically odd, and could evoke additional processing efforts because of integration difficulties. We hypothesized that this would be reflected by a P600 effect, which was confirmed by the ERP results. However, we found a similar effect for the incoherent words which were preceded by *eigenlijk*, and we subsequently argued that the P600 effect could reflect pragmatic inferencing which belongs to the domain of social cognition, as the comprehender tries to uncover the attitude and intention of the speaker. This intensive pragmatic processing seems to take place regardless of the coherence between the information encountered and the discourse context. Further experimental investigations are needed to verify the exact nature of such pragmatic inferences.

To sum up, we started out to investigate whether *eigenlijk* affects predictive processing in terms of pre-activation and integration of unexpected information. In the present study we have not found any evidence for such a facilitating effect of *eigenlijk*. However, even though *eigenlijk* may not directly influence the lexical-semantic processing of words, it does play a role for the composition of the discourse representation in terms of socio-pragmatic inferences. The present study thus reminds us once again that language is not just about information transmission, and that comprehenders do not merely use words as cues to facilitate the updating of their discourse model in terms of propositional content. We do also use language to sustain our relationships with others, and specific words to communicate our intentions and attitude – which comprehenders reflect upon through effortful processing.

References

Aijmer, K., & Simon-Vandenbergen, A. M. (2004). A model and a methodology for the study of pragmatic markers: the semantic field of expectation. *Journal of Pragmatics*, *36*(10), 1781–1805.

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

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*(3), 255–278.

Bašnáková, J., Van Berkum, J., Weber, K., & Hagoort, P. (2015). A job interview in the MRI scanner: How does indirectness affect addressees and overhearers? *Neuropsychologia*, *76*, 79–91.

Bašnáková, J., Weber, K., Petersson, K. M., Van Berkum, J., & Hagoort, P. (2014). Beyond the Language Given: The Neural Correlates of Inferring Speaker Meaning. *Cerebral Cortex*, *24*(10), 2572–2578.

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never Say No … How the Brain Interprets the Pregnant Pause in Conversation. *PLOS ONE*, *10*(12), e0145474.

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native “um”s elicit prediction of low-frequency referents, but non-native “um”s do not. *Journal of Memory and Language*, *75*, 104–116.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Neuroreport*, *18*(17), 1851–1854.

Canestrelli, A., & Mak, W. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, *28*(9), 1394–1413.

Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.

Corley, M., & Hartsuiker, R. (2003). Hesitation in speech can... um... help a listener understand. In *Proceedings of the 25th meeting of the cognitive science society* (pp. 276–281).

Corley, M., MacGregor, L., & Donaldson, D. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*, 658–668.

Coulson, S., & Kutas, M. (2001). Getting it: Human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters*, *316*(2), 71–74.

Coulson, S., & Lovett, C. (2004). Handedness, hemispheric asymmetries, and joke comprehension. *Cognitive Brain Research*, *19*, 275–288.

Coulson, S., & Lovett, C. (2010). Comprehension of non-conventional indirect requests: An event-related brain potential study. *Italian Journal of Linguistics*, *22*(1), 107–124.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-Processing in Sentence Comprehension: Sensitivity to Likely Upcoming Meaning and Structure. *Language and Linguistics Compass*, *8*(12), 631–645.

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*(9), 1203–1207.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121.

DeLong, K., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162.

DeLong, K., Urbach, T., & Kutas, M. (2017). *Concerns with Nieuwland et al. (2017)*. Retrieved from http://kutaslab.ucsd.edu/pdfs/FinalDUK17Comment9LabStudy.pdf

Diewald, G. (2010). On some problem areas in grammaticalization studies. In K. Stathi, E. Gehweiler, & E. König (Eds.), *Grammaticalization: Current views and issues* (pp. 17–50). Amsterdam/Philadelphia: John Benjamins Publishing.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84.

Federmeier, K., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, *31*, 931–952.

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLoS ONE*, *10*(3), e0120068.

Grondelaers, S., Speelman, D., Drieghe, D., Brysbaert, M., & Geeraerts, D. (2009). Introducing a new entity into discourse: Comprehension and production evidence for the status of Dutch er "there" as a higher-level expectancy monitor. *Acta Psychologica*, *130*(2), 153–160.

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

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*(1), 59–73.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225.

Kolk, H. H. J., Chwilla, D. J., Van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*, 1–36.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 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. *Language and Cognitive Processes*, *21*(5), 489–530.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.

Kuperberg, G., Sitnikova, T., Caplan, D., & Holcomb, P. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117–129.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*(4), 533–572.

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*, 621–47.

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). Package "lmerTest."

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 Effects of Prediction from Association in Single-word Contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502.

Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., & Levinson, S. C. (2014). Early Anticipation Lies behind the Speed of Response in Conversation. *Journal of Cognitive Neuroscience*, *26*(11), 2530–2539.

Maschler, Y., & Schiffrin, D. (2015). Discourse markers: Language, meaning, and context. In D. Tannen, H. Hamilton, & D. Schiffrin (Eds.), *The Handbook of Discourse Analysis* (2nd ed., pp. 189–221). Chichester, UK: John Wiley & Sons.

Moreno, E., Federmeier, K., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*, *80*(2), 188–207.

Nieuwland, M., & Martin, A. (2012). If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition*, *122*, 102–109.

Nieuwland, M., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2017). Limits on prediction in language comprehension: A multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. *BioRxiv*, 111807.

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is too hard to handle. *Psychological Science*, *19*(12), 1213–1218.

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

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't It Ironic? An Electrophysiological Exploration of Figurative Language Processing. *Journal of Cognitive Neuroscience*, *23*(2), 277–293.

Schiffrin, D. (1988). *Discourse markers*. Cambridge: Cambridge University Press.

Spotorno, N., Cheylus, A., Van Der Henst, J. B., & Noveck, I. A. (2013). What's behind a P600? Integration Operations during Irony Processing. *PLoS ONE*, *8*(6).

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*, 382–392.

Van Bergen, G. (2017). *Expectation management in conversation: interpersonal discourse particles modulate discourse-based predictions in incremental language comprehension*. Manuscript submitted for publication.

Van Bergen, G., & Hogeweg, L. (2017). Managing alignment in dialogue: The role of contrastive discourse particles in Dutch. Manuscript submitted for publication.

Van Bergen, G., Van Gijn, R., Hogeweg, L., & Lestrade, S. (2011). Discourse marking and the subtle art of mind - reading : The case of Dutch eigenlijk. *Journal in Pragmatics*, *43*, 3877–3892.

Van Bergen, G., Van Gijn, R., Hogeweg, L., & Lestrade, S. (2011). Discourse marking and the subtle art of mind-reading: The case of Dutch eigenlijk. *Journal of Pragmatics*, *43*(15), 3877–3892.

Van Berkum, J. (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., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467.

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic Integration in Sentences and Discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657–671.

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 Berkum, J., Van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*, 580–591.

Van de Meerendonk, N., Kolk, H. H., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Linguistics and Language Compass*.

Van Petten, & Luka, B. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176–190.

Wicha, N. Y. 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. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288.

Wilson, D. (2009). Irony and metarepresentation. *UCL Working Papers in Linguistics*, *21*, 183–226.

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*(6), 648–672.

Yan, S., & Kuperberg, G. R. (2017). *Prediction (or not) during language processing. A commentary on Nieuwland et al. (2017) and DeLong et al. (2005)*. *bioRxiv*. Retrieved from http://biorxiv.org/content/early/2017/05/30/143750.abstract

Appendix A – Selection criteria stimuli

The selection of the 144 sets out of the total 180 sets that have been pretested (see section 4) was based on the cloze probabilities at the Discourse level and the following criteria:

(1) The cloze probabilities of the Neutral and Eigenlijk condition differ in the expected direction:
   a. Cloze probability Discourse-Target: Neutral ≥ Eigenlijk
   b. Cloze probability Discourse-Competitor: Neutral ≤ Eigenlijk

(2) The Neutral condition should be constraining enough towards the Target, and not evoke too many Competitor responses:
   a. Cloze probability Discourse-Target: Neutral ≥ 0.35
   b. Cloze probability Discourse-Competitor: Neutral ≤ 0.45

(3) The "*Eigenlijk* manipulation" should be successful.
   a. Cloze probability Discourse-Target: Eigenlijk ≤ 0.80
   b. Cloze probability Discourse-Competitor: Eigenlijk ≥ 0.25

Sets were excluded if they did not meet these criteria. Besides, two additional sets have been excluded because it was not possible to objectively interpret and code the responses.

Appendix B – Mixed-effects models results (pretest)

Table B. Mixed-effects models results for the four coded categories of the discourse completion test, with the conditions Eigenlijk, Inderdaad and Neutral as factors (with Neutral as intercept).

TargetLemma		**ALL**			**SUBSET**		
Random factors		Variance	*SD*		Variance	*SD*	
Item	(Intercept)	1.422	1.192		1.265	1.125	
	Eigenlijk	0.749	0.866		0.718	0.847	
	Inderdaad	0.231	0.481		0.195	0.442	
Subject	(Intercept)	0.124	0.352		0.129	0.359	
	Eigenlijk	0.484	0.696		0.521	0.722	
	Inderdaad	0.096	0.310		0.080	0.283	
Fixed factors		***β***	***SE***	***P***	***β***	***SE***	***p***
Intercept		-0.052	0.107	0.629	-0.016	0.113	0.887
Eigenlijk		-2.331	0.140	< .001	-2.455	0.152	< .001
Inderdaad		0.431	0.077	< .001	0.315	0.080	< .001

TargetDiscourse		**ALL**			**SUBSET**		
Random factors		Variance	*SD*		Variance	*SD*	
Item	(Intercept)	1.763	1.328		1.168	1.081	
	Eigenlijk	1.305	1.142		0.934	0.966	
	Inderdaad	1.870	1.368		1.352	1.163	
Subject	(Intercept)	0.412	0.642		0.504	0.710	
	Eigenlijk	0.817	0.904		0.894	0.945	
	Inderdaad	0.331	0.575		0.322	0.567	
Fixed factors		***β***	***SE***	***p***	***β***	***SE***	***p***
Intercept		1.687	0.143	< .001	1.831	0.145	< .001
Eigenlijk		-3.300	0.168	< .001	-3.505	0.174	< .001
Inderdaad		2.184	0.222	< .001	2.009	0.236	< .001

CompLemma		**ALL**			**SUBSET**		
Random factors		Variance	*SD*		Variance	*SD*	
Item	(Intercept)	1.657	1.287		1.670	1.292	
	Eigenlijk	0.484	0.696		0.450	0.671	
	Inderdaad	19.94	4.47		23.306	4.828	
Subject	(Intercept)	0.208	0.456		0.238	0.488	
	Eigenlijk	0.107	0.328		0.031	0.176	
	Inderdaad	0.590	0.768		0.370	0.608	
Fixed factors		***β***	***SE***	***p***	***β***	***SE***	***p***
Intercept		-3.505	0.179	< .001	-3.645	0.206	< .001
Eigenlijk		2.312	0.160	< .001	2.470	0.182	< .001
Inderdaad		-5.726	1.341	< .001	-6.046	1.583	< .001

CompDiscourse		**ALL**			**SUBSET**		
Random factors		Variance	*SD*		Variance	*SD*	
Item	(Intercept)	1.624	1.274		1.153	1.074	
	Eigenlijk	1.171	1.082		0.905	0.951	
	Inderdaad	2.063	1.436		1.722	1.312	
Subject	(Intercept)	0.479	0.692		0.516	0.719	
	Eigenlijk	0.775	0.880		0.850	0.922	
	Inderdaad	0.366	0.605		0.204	0.452	
Fixed factors		***β***	***SE***	***p***	***β***	***SE***	***p***
Intercept		-2.149	0.149	< .001	-2.27	0.153	< .001
Eigenlijk		3.403	0.165	< .001	3.622	0.175	< .001
Inderdaad		-2.650	0.315	< .001	-2.543	0.337	< .001

Appendix C – Mixed-effects models structures (pretest)

```
cloze.glmm.targetlemma = glmer(TargetLemma ~ Cond + (1+Cond|SetNr) +
(1+Cond|UserId), data=cloze_select,  family="binomial",
glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 100000)))
```

```
cloze.glmm.targetdiscourse = glmer(TargetDiscourse ~ Cond + (1+Cond|SetNr) +
(1+Cond|UserId), data=cloze_select, family="binomial",
glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 100000)))
```

```
cloze.glmm.complemma = glmer(CompLemma ~ Cond + (1+Cond|SetNr) + (1+Cond|UserId),
data=cloze_select, family="binomial", glmerControl(optimizer="bobyqa", optCtrl =
list(maxfun = 100000)))
```

```
cloze.glmm.compdiscourse = glmer(CompDiscourse ~ Cond + (1+Cond|SetNr) +
(1+Cond|UserId), data=cloze_select, family="binomial",
glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 100000)))
```

Appendix D – Grand-averaged ERP waveforms

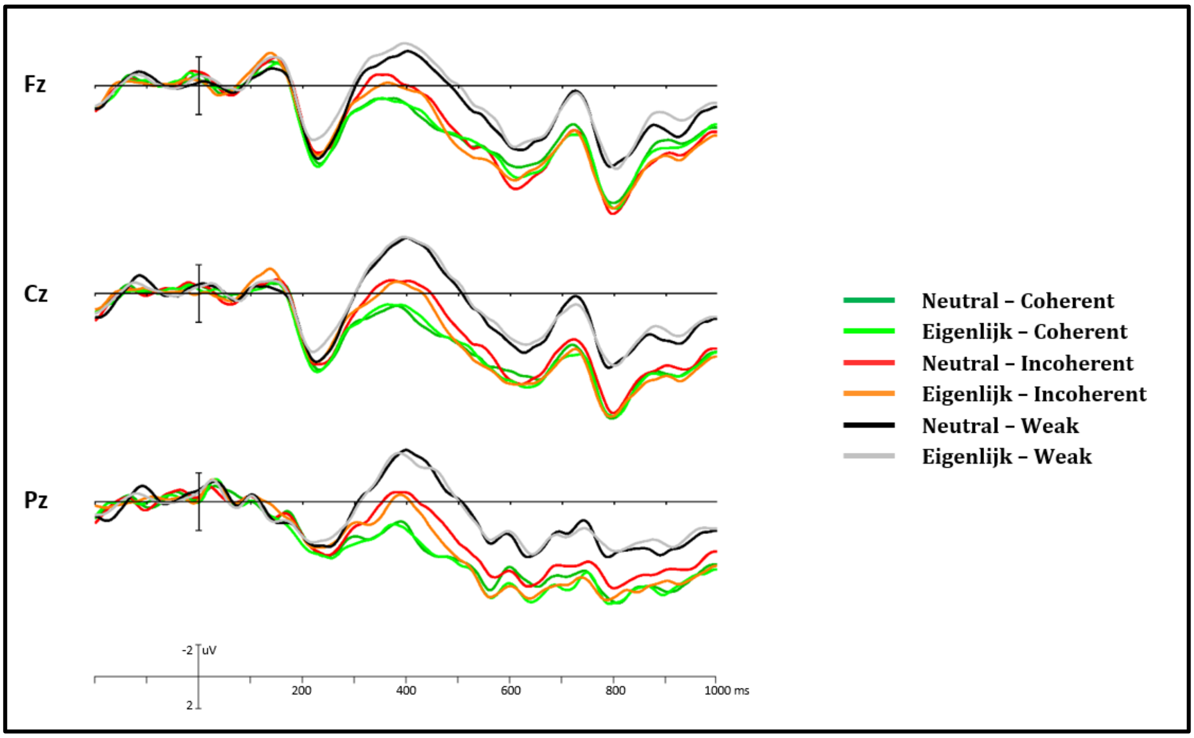

Figure D. Grand-averaged (N = 27) waveforms (negative plotted up) to critical words for electrodes Fz, Cz and PZ. Apart from the four experimental conditions discussed in this rapport, the two additional weak-constraining conditions are plotted (Weak-Neutral and Weak-Eigenlijk).