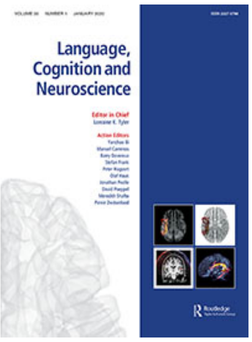

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Anticipating predictability: an ERP investigation of expectation-managing discourse markers in dialogue comprehension

Marlou Rasenberg, Joost Rommers & Geertje van Bergen

LANGUAGE, COGNITION AND NEUROSCIENCE
2020, VOL. 35, NO. 1, 1–16
https://doi.org/10.1080/23273798.2019.1624789

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS Check for updates

Anticipating predictability: an ERP investigation of expectation-managing discourse markers in dialogue comprehension

Marlou Rasenberg[a,b], Joost Rommers[c] and Geertje van Bergen[b,c]

[a]Centre for Language Studies, Radboud University, Nijmegen, Netherlands; [b]Neurobiology of Language, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; [c]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

ABSTRACT
In two ERP experiments, we investigated how the Dutch discourse markers *eigenlijk* "actually", signalling expectation disconfirmation, and *inderdaad* "indeed", signalling expectation confirmation, affect incremental dialogue comprehension. We investigated their effects on the processing of subsequent (un)predictable words, and on the quality of word representations in memory. Participants read dialogues with (un)predictable endings that followed a discourse marker (*eigenlijk* in Experiment 1, *inderdaad* in Experiment 2) or a control adverb. We found no strong evidence that discourse markers modulated online predictability effects elicited by subsequently read words. However, words following *eigenlijk* elicited an enhanced posterior post-N400 positivity compared with words following an adverb regardless of their predictability, potentially reflecting increased processing costs associated with pragmatically driven discourse updating. No effects of *inderdaad* were found on online processing, but *inderdaad* seemed to influence memory for (un)predictable dialogue endings. These findings nuance our understanding of how pragmatic markers affect incremental language comprehension.

ARTICLE HISTORY
Received 16 July 2018
Accepted 20 May 2019

KEYWORDS
Discourse processing; discourse markers; dialogue; EEG; predictability; pragmatics

Introduction

Language comprehenders seem to use the available context to actively predict upcoming linguistic information (for reviews, see e.g. Federmeier, 2007; Kamide, 2008; Kuperberg & Jaeger, 2016; Kutas, DeLong, & Smith, 2011; Pickering & Garrod, 2013; Van Petten & Luka, 2012). Predictability effects in language processing have been particularly well-established in event-related brain potentials (ERPs), which can provide functionally specific measures of the cognitive and neural processes involved in incremental dialogue comprehension. However, it remains unclear how the broad range of informative lexical, syntactic and pragmatic cues in the context is used to generate and revise ongoing predictions. Against this background, the current study examined how previously established ERP effects of predictability are modulated by two specific pragmatic cues: Dutch *eigenlijk* (≈ "actually, in fact") and *inderdaad* (≈ "indeed").

Inderdaad and *eigenlijk* belong to the broad class of discourse markers (also referred to as pragmatic markers, pragmatic particles, discourse particles, or discourse connectives), which are linguistic elements that encode a relation between the sentence in which they occur and the surrounding discourse situation (e.g. Aijmer, 2002; Fischer, 2006; Fraser, 1999; Schiffrin, 1988). By using *inderdaad* and *eigenlijk*, speakers can demonstrate sensitivity to their addressee's likely expectations, as illustrated in the following constructed Dutch dialogue:

(1) A: *Je hebt vast genoten van het New York Philharmonisch Orkest?*
You must have enjoyed the New York Philharmonic Orchestra?
B: *Ik vond hun concert inderdaad prachtig.*
I found their concert [indeed] beautiful.
B′: *Ik vond hun concert eigenlijk verschrikkelijk.*
I found their concert [actually] horrible.

In the answers in (1), *inderdaad* and *eigenlijk* respond to the expectation that can be inferred from A's suggestive question: they mark either alignment (*inderdaad*) or misalignment (*eigenlijk*) between what B says and what B thinks A expects to hear.

Theoretically, discourse markers are assumed to manage the course of the conversation: they "function as instructions from the speaker to the hearer on how to integrate the host unit into a coherent mental representation of the discourse" (Mosegaard-Hansen, 1998,

CONTACT Geertje van Bergen Geertje.vanBergen@mpi.nl
Supplemental data for this article can be accessed https://doi.org/10.1080/23273798.2019.1624789.

p. 358; see also Aijmer & Simon-Vandenbergen, 2004; Blakemore, 2002; Fox Tree, 2010; Schourup, 1999). Whereas theoretical research on discourse markers has extensively investigated under which conditions they are used by the speaker, surprisingly little is known about how they are used by the comprehender. Although there are studies on temporal and causal discourse connectives (e.g. Canestrelli, Mak, & Sanders, 2013; Nieuwland, 2015; Xiang & Kuperberg, 2015) and focus particles (Gerwien & Rudka, in press; Kim, Gunlogson, Tanenhaus, & Runner, 2015), these have not been investigated in conversational contexts, which is the typical environment of *eigenlijk* and *inderdaad*. Effects of disfluencies (*uh/uhm*) and repairs (*oh, I mean*) on language processing have been studied in interactive contexts (e.g. Fox Tree, 2001; Fox Tree & Schrock, 1999), but their assumed conversation-managing function (Clark & Fox Tree, 2002) remains controversial (e.g. Finlayson & Corley, 2012; Schegloff, 2010).

The present study empirically investigated the theoretically assumed function of *eigenlijk* and *inderdaad* by examining their effects on online language comprehension. We investigated to what extent comprehenders use the pragmatic information encoded in *eigenlijk* and *inderdaad* to guide their expectations about likely dialogue continuations during reading. This allowed us to refine theoretical claims about the facilitating role of discourse markers for the comprehender, which are almost exclusively based on language production and hence remain underspecified regarding the affected comprehension processes.

Electrophysiological effects of word predictability

We assessed multiple electrophysiological signatures of predictability-related processing to address whether, and at which processing stages, the presence of *eigenlijk* or *inderdaad* affected processing of subsequent (un)predictable words. One focus was the N400, a negativity that peaks over centro-parietal sites around 400 ms after the onset of a potentially meaningful stimulus. Although some disagreements remain regarding this component's exact interpretation in terms of retrieval or integration (Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Brown & Hagoort, 1993; Kutas & Federmeier, 2000, 2011; Nieuwland et al., in press; van Berkum, 2009), there is broad consensus that the N400 reflects semantic processing. Its amplitude is strongly negatively correlated with a word's cloze probability in a sentence (Kutas & Hillyard, 1980, 1984), that is, the proportion of participants who complete a truncated version of the sentence with that word in an offline task. N400 amplitude is also reduced by predictability based on extra-sentential information, such as the wider discourse context (e.g. Federmeier & Kutas, 1999; Nieuwland & van Berkum, 2006; Otten & van Berkum, 2008), specific pragmatic expressions (e.g. negation, Nieuwland & Kuperberg, 2008; scalar statements, Nieuwland, Ditman, & Kuperberg, 2010; counterfactuals, Nieuwland & Martin, 2012; connectives, Xiang & Kuperberg, 2015), general world knowledge (e.g. Hagoort, Hald, Bastiaansen, & Petersson, 2004), and voice-based pragmatic inferences about speaker characteristics (van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008). In addition, N400 amplitude is sensitive to long-term memory structure, in that it is attenuated in response to words that are unpredictable, if they are semantically related to predictable words (Federmeier & Kutas, 1999).

Several ERP studies have also reported post-N400 positivity effects related to predictability. Unlike the N400, which reflects the benefits of a supportive context for semantic processing, such late positivity effects are thought to reflect additional processing costs associated with some aspect of prediction disconfirmation. Frontally distributed effects have been reported in response to words that are unpredictable but plausible sentence continuations (e.g. DeLong, Quante, & Kutas, 2014; DeLong, Urbach, Groppe, & Kutas, 2011; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Moreno, Federmeier, & Kutas, 2002; Thornhill & Van Petten, 2012; Van Petten & Luka, 2012). Posteriorly distributed post-N400 positivity effects (sometimes called P600 or late positive complex (LPC) effects) are typically elicited in response to anomalous input (e.g. DeLong et al., 2014; Kuperberg, 2007; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; van de Meerendonk, Kolk, Vissers, & Chwilla, 2010), although similar effects have also been reported in response to non-anomalous pragmatic phenomena, such as metonymy (Schumacher, 2013), jokes as opposed to non-funny control stimuli (Coulson & Lovett, 2004), indirect requests relative to literal statements (Coulson & Lovett, 2010) and ironic as opposed to literal sentences (Regel, Gunter, & Friederici, 2011; Spotorno, Cheylus, van der Henst, & Noveck, 2013). Such posterior brain responses have been associated with some form of combinatorial processing, reflecting reanalysis, reorganisation or updating of the discourse model (e.g. Brouwer, Fitz, & Hoeks, 2012; Kuperberg, 2007; Schumacher, 2013; van de Meerendonk et al., 2010). In sum, given their sensitivity to both predictability and pragmatic factors, the N400 and frontal and posterior post-N400 positivities provide relevant windows into multiple aspects of processing that expectation-managing discourse markers could modulate.

The present study

In the present study, participants read dialogues (see Table 1) for comprehension while their EEG was recorded. The dialogues ended in contextually predictable or less predictable words, and the presence of the discourse markers *eigenlijk* (Experiment 1) or *inderdaad* (Experiment 2) was manipulated. Dialogues containing a discourse marker were compared with the same dialogues in which the discourse marker was replaced by a control adverb (e.g. *gisteren* "yesterday"; see Methods). We evaluated the above-mentioned ERP components to address whether, and at which processing stages, the presence of *inderdaad* and *eigenlijk* affected processing of subsequent (un)predictable words.

In control dialogues with an adverb, we expected reduced N400 amplitudes in response to predictable relative to unpredictable dialogue continuations, in line with prior research. We hypothesised that, if discourse markers affect semantic processing of subsequently read input, then having encountered *inderdaad* (as opposed to an adverb) should result in a further reduction of N400 amplitude elicited by predictable words, and possibly increased N400 amplitude in response to unpredictable words. Conversely, we expected that encountering the warning signal *eigenlijk* (relative to the control adverb) would yield reduced N400 amplitude in response to unpredictable words, and possibly enhanced N400 amplitude elicited by predictable words. Furthermore, if pragmatic expectation-managing cues impact later processing stages, we predicted *eigenlijk*- and *inderdaad*-based modulations of post-N400 positivities. In line with previous studies, we expected a frontally distributed post-N400 positivity in response to unpredictable but plausible words, relative to predictable words, in control dialogues without a discourse marker. This predictability effect was hypothesised to be enhanced in response to words following a confirmatory cue (*inderdaad*), and attenuated or reversed for words following an adversative cue (*eigenlijk*). In addition, the pragmatic anomaly created by presenting less predictable words after *inderdaad* (normally associated with expectation confirmation), or predictable words after *eigenlijk* (normally marking unexpectedness), could lead to integration difficulty and hence elicit a posterior post-N400 positivity relative to pragmatically congruent conditions (e.g. DeLong et al., 2014).

Finally, after reading the dialogues, participants performed a recognition memory task, which allowed us to explore whether the presence of expectation-managing discourse markers affected the quality of the representations of predictable and unpredictable words in memory. Previous research has investigated how word predictability affects memory, but without manipulating discourse markers. Some studies have shown better memory for predictable words than for less predictable words (e.g. Miller & Selfridge, 1950; Riggs, Wingfield, & Tun, 1993), perhaps due to a better integrated memory trace. Other studies have shown poorer memory for predictable words than for unpredictable words (e.g. Cairns, Cowart, & Jablon, 1981; Corley, MacGregor, & Donaldson, 2007; Federmeier et al., 2007; O'Brien & Myers, 1985; Perry & Wingfield, 1994), possibly because more predictable input is processed less thoroughly (e.g. Rommers & Federmeier, 2018a; van Berkum, 2010). We hypothesised

Table 1. Example of an experimental stimulus in each condition.

Context:	*Diana is een weekend met haar klas van de kunstacademie naar Parijs geweest.* Diana spent a weekend in Paris with her art academy class.		
Question:	*Haar vriendin vraagt: jullie hebben vast een hoop kunst gezien?* Her friend asks: you guys must have seen a lot of art?		
Answer:	*Diane zegt: …* Diane says: …		
Condition	Plain *Predictability*	Pragmatic *Coherence*	*Target sentence*
Predictable Adverb	Predictable	n/a	*We zijn daar elke dag naar een **museum** geweest.* We have there every day to a museum been.
Unpredictable Adverb	Unpredictable	n/a	*We zijn daar elke dag naar een **park** geweest.* We have there every day to a park been.
Experiment 1			
Predictable *Eigenlijk*	Predictable	Incoherent	*We zijn eigenlijk elke dag naar een **museum*** geweest. *We have [actually] every day to a museum been.*
Unpredictable *Eigenlijk*	Unpredictable	Coherent	*We zijn eigenlijk elke dag naar een **park** geweest.* We have [actually] every day to a park been.
Experiment 2			
Predictable *Inderdaad*	Predictable	Coherent	*We zijn inderdaad elke dag naar een **museum*** geweest. We have [indeed] every day to a museum been.
Unpredictable *Inderdaad*	Unpredictable	Incoherent	*We zijn inderdaad elke dag naar een **park** geweest.* We have [indeed] every day to a park been.

Notes: Dutch word order was maintained in the translations of the target sentences. The critical manipulations are underlined (control adverb/discourse marker) or in boldface (critical word; CW). "Plain Predictability" refers to the predictability of the CW in the adverb conditions (i.e. predictability in the absence of a discourse marker); "Pragmatic Coherence" indicates the pragmatic fit of the (un)predictable CW when following a discourse marker.

that potential memory differences between predictable and unpredictable words would be enhanced for words following a pragmatic prediction confirmation cue (*inderdaad*), and reduced for words following a pragmatic prediction disconfirmation cue (*eigenlijk*).

Materials and methods

Participants

Each experiment was conducted with 40 participants (Exp. 1: 27 female, mean age 22.1 years, range 18–28 years; Exp. 2: 32 female, mean age 22.5 years, range 19–35 years) from the Max Planck Institute for Psycholinguistics participant pool. All were native speakers of Dutch and right-handed. Participants provided informed consent prior to starting the experiment and were paid 18 euros for participation. The study received ethical approval from the Faculty of Social Sciences at Radboud University Nijmegen (#ECG2013-1308-120).

Data from three participants in Experiment 1 and six participants in Experiment 2 that exhibited excessive artifacts such as blinks, drifts, eye movements or excessive muscle activity (>30% of trials affected) were excluded from further analyses. Two additional participants were excluded from Experiment 1 because of poor performance in the memory task (negative discriminability values, likely due to a misinterpretation of the response scale or a lack of attention during the reading experiment). Hence, for Experiment 1, 35 participants were included in the ERP analyses and 38 in the behavioural memory analyses; in Experiment 2, ERP analyses included 34 participants and behavioural memory analyses were performed on all 40 participants.

Materials

Experimental items consisted of 144 Dutch written conversations in easily imaginable situations. Each item consisted of an introductory context sentence, followed by a question-answer pair. The combination of the context sentence and the question were created to evoke a specific lexical-semantic prediction. Answers occurred in four conditions that combined two factors: a) the answers contained a critical word that was either predictable or unpredictable on the basis of the prior context, and b) the critical words were either preceded by an adverb or adverbial phrase (e.g. *gisteren, "*yesterday", *graag*, "happily"; adverbs varied across items and were included to keep sentence structure comparable across conditions) or by an expectation-managing discourse marker (*eigenlijk* in Experiment 1, *inderdaad* in Experiment 2). An example of an experimental item in all conditions is presented in Table 1.

Predictability of the critical word was determined on the basis of a web-based cloze test, in which participants read 180 experimental conversations in three conditions (adverb vs. *eigenlijk* vs. *inderdaad*; 60 items per condition, counterbalanced across three lists; 20 participants per list), and were asked to finish the truncated answer. We selected 72 lemmas with the highest cloze probability in the Adverb-condition as predictable critical words. The 72 unpredictable critical words were selected from the completions provided in the *Eigenlijk*-condition to ensure their semantic plausibility; these words had a lower cloze probability in the baseline condition. Note that the denotation of critical words as predictable or unpredictable follows from their cloze probabilities in the Adverb-conditions (referred to as *Plain Predictability* in Table 1), rather than from their cloze values in the discourse marker conditions. For example, a critical word with high cloze probability in the *Inderdaad*-condition but low cloze probability in the *Eigenlijk*-condition will have high *Plain Predictability* in both cases. Mean cloze probabilities of the critical words in the 144 selected items in each condition are provided in Table 2.

A logistic mixed-effects regression analysis of cloze probabilities of the selected critical words showed a Predictability by Discourse Marker cross-over interaction (comparing models with vs. without the interaction effect: χ^2 (2) = 2616, $p < 0.001$). Follow-up analyses of predictable and unpredictable critical words separately showed that, relative to the Adverb-condition, cloze probabilities of plain-predictable critical words were

Table 2. Cloze probabilities of critical words (CWs) in each condition.

Condition	Plain Predictability	Pragmatic Coherence	CW cloze *M*	CW cloze [SD]
Predictable Adverb	Predictable	n/a	0.49	[0.50]
Unpredictable Adverb	Unpredictable	n/a	0.05	[0.23]
Experiment 1				
Predictable *Eigenlijk*	Predictable	Incoherent	0.15	[0.36]
Unpredictable *Eigenlijk*	Unpredictable	Coherent	0.27	[0.45]
Experiment 2				
Predictable *Inderdaad*	Predictable	Coherent	0.55	[0.50]
Unpredictable *Inderdaad*	Unpredictable	Incoherent	0.01	[0.07]

higher in the *Inderdaad*-condition ($\beta = 0.32$, SE = 0.08, $p < 0.001$) and lower in the *Eigenlijk*-condition ($\beta = -2.46$, SE = 0.15, $p < 0.001$); conversely, cloze probabilities of plain-unpredictable critical words were lower in the *Inderdaad*-condition ($\beta = -6.05$, SE = 1.58, $p < 0.001$) and higher in the *Eigenlijk*-condition ($\beta = 2.47$, SE = 0.18, $p < 0.001$) when compared with the Adverb-condition. Predictable and unpredictable critical words were comparable in terms of length ($M = 7.0$ vs. 6.7 characters, respectively) and frequency ($M = 67$ vs. 85 per million words in CELEX; Baayen, Piepenbrock, & Gulikers, 1995).

Dialogues in each condition were divided across four counterbalanced lists such that participants would see each experimental item in only one condition. Seventy-two filler items were added to each list. Filler items had the same structure as the experimental conversations (context sentence followed by a question-answer pair) and contained an adverb or a discourse marker (*eigenlijk* in Experiment 1, *inderdaad* in Experiment 2), but all contexts were weakly constraining.[1] In sum, the resulting lists each consisted of 216 conversations, half of which contained an expectation-managing discourse marker in the answer. Lists were pseudo-randomized individually for each participant.

For the memory test, a subset of 96 critical words were selected under the constraint that they appeared only once in the experimental conversations (24 per condition). Each list was supplemented with 48 words that occurred in the filler conversations, and 48 new (unseen) words similar in length and frequency to the old (seen) words. This resulted in 192 words per list; the order of items on each list was pseudo-randomized for each participant.

Procedure

Participants were tested individually in a dimly lit soundproof booth, seated at a viewing distance of approximately 100 cm from a computer screen. For the reading phase, participants were instructed to silently and attentively read the conversations for comprehension, and to avoid blinks, muscle movements and eye movements.

Stimuli were presented in a black Lucida Console font, 26-point size, on a white background. Each trial started with a centred fixation cross which remained on the screen for 1000 ms, followed by a 500 ms blank screen. Next, the context sentence was presented in full, followed by the question that also occurred in full. Participants read both sentences at their own pace, and pressed a button to continue. After the question, the first part of the answer (e.g. "Diane says") appeared and remained on the screen for 800 ms. The presentation time of the subsequent words in the answer sentence was variable to mimic relatively natural reading (e.g. Nieuwland & van Berkum, 2006). Word duration was computed as (number of letters * 30) + 190 ms, with a maximum of 400 ms. Adverbs/discourse markers and critical words had a fixed duration of 400 ms and the inter-stimulus interval was fixed at 150 ms. The final word of the sentence appeared with a period and was presented for 800 ms, after which the next trial started automatically. Participants started with 4 practice items, and then completed 6 blocks of 36 experimental items, separated by self-timed breaks. This part of the experiment took 45–60 min.

After the reading experiment, participants took a 30 s math test to clear their verbal short-term memory, after which they started the word recognition test. Participants were asked to judge whether they recognised words from the reading experiment, indicating how confident they were of their response (*zeker nieuw* "sure new"; *misschien nieuw* "maybe new"; *misschien oud* "maybe old"; *zeker oud* "sure old"). Each target word (black Lucida Console font, 26-point size) was presented in the centre of the screen. After 1500 ms, the four answer options appeared below the target word, matching in colour and linear order with four buttons on the button box. Participants were instructed to wait for the answer options to appear on the screen before pressing a button. It took participants about 15 min to complete the memory test; the full experimental session took on average 2 h.

EEG recording and preprocessing

The EEG was recorded from 31 active cap-mounted Ag/AgCl electrodes (actiCAP, Brain Products GmbH), referenced online to the left mastoid. Blinks and eye movements (EOG) were measured via four electrodes located at the outer canthi of both eyes and above and below the left eye. Electrode impedances were kept below 20 kΩ. Signals were amplified using BrainAmp DC amplifiers with a band-pass filter between 0.01 and 150 Hz and digitised at a sampling frequency of 500 Hz. The EEG signal was re-referenced to the mean of the left and right mastoids and bipolar EOG derivations were created. The continuous EEG was filtered with a 0.1 Hz high-pass filter (two-pass Butterworth with a 12 dB/oct roll-off) and segmented into epochs encompassing the signal from −200 ms until 1000 ms relative to the onset of the critical word. A 200 ms pre-stimulus baseline was subtracted. Segments containing blinks, drifts, eye movements, or excessive muscle activity were removed in a semi-automatic fashion using participant-specific thresholds. In Experiment 1, a total of 9% of the trials was removed, with similar trial numbers remaining across conditions: Predictable Adverb 33 ± 3 (mean ± SD), Predictable Eigenlijk 33 ± 3, Unpredictable Adverb 33 ± 2, Unpredictable

Eigenlijk 33 ± 2. In Experiment 2, the overall trial loss was 8%; similar trial numbers across conditions remained: Predictable Adverb 33 ± 3, Predictable Inderdaad 33 ± 2, Unpredictable Adverb 33 ± 3, Unpredictable Inderdaad 33 ± 3.

Event-related potentials

Trials were averaged in the time domain for each condition and each participant, forming ERPs. A 20 Hz low-pass filter was applied (two-pass Butterworth with a 24 dB/oct roll-off). Mean amplitudes were computed for two pre-specified time windows: 300-500 ms (N400) and 500-800 ms (post-N400 positivity), averaged across channels. N400 mean amplitude was measured over centro-parietal sites, where the effect is usually maximal (Cz, C3, C4, CP1, CP2, Pz). Post-N400 mean amplitude was measured across 10 frontal-central channels (Fp1, Fp2, F3, F4, F7, F8, FC1, FC2, FC5, FC6)[2] to capture anterior positivity effects (similar to Federmeier et al., 2007), and across 10 parietal-occipital channels (CP1, CP2, CP5, CP6, P3, P4, P7, P8, O1, O2) to capture posterior positivity effects (similar to Kuperberg et al., 2006).

Data analysis

In correspondence with most of the literature, ANOVAs are reported, supplemented with confidence intervals and effect sizes (Cohen's d_z for within-subject designs). ANOVAs included the factors Plain Predictability (predictable, unpredictable), DM (adverb, *eigenlijk/inderdaad*) and their interaction as within-subject variables. In order to investigate the topographic distribution of the post-N400 positivity effects, Hemisphere was included as additional variable.

EEG data acquired during the memory test were not further analyzed; after data preprocessing too few trials were left per condition. Behavioural memory performance was analyzed by computing a standard signal detection-theoretic index of discriminability (d_a) per condition per participant. Discriminability was analyzed by means of repeated-measures ANOVAs including Predictability, Discourse Marker, and their interaction.

Results

Experiment 1 (eigenlijk)

Event-related potentials

Figure 1 shows the ERPs time-locked to the presentation of the critical words. After a visual P1, N1 and P2, an N400 was elicited, followed by a late positive-going wave. Figure 2 presents the scalp topographies of the difference waves between conditions in the N400 and post-N400 positivity time window.

Corroborating numerous previous findings, N400 amplitude in response to predictable words was attenuated by 1.48 µV (95% CI [1.24, 1.71], $d_z = 0.56$) compared with the N400 to unpredictable words, F (1, 34) = 35.817, $p < .0001$. We had hypothesised that the presence of *eigenlijk* would attenuate or reverse this plain predictability effect, but we found no evidence for a Predictability x Discourse Marker interaction, F (1, 34) = 1.544, $p = 0.225$, nor did we find a main effect of Discourse Marker, F (1, 34) = 0.824, $p = 0.370$.

With respect to later processing stages, we hypothesised a frontally distributed post-N400 positivity effect for unpredictable but plausible dialogue continuations (e.g. DeLong et al., 2014). Surprisingly, analyses of mean amplitudes in the post-N400 time window revealed no evidence for a difference in anterior post-N400 positivity between predictable and unpredictable words, F (1, 34) = 1.567, $p = 0.219$. Neither did we find evidence for a difference between words following *eigenlijk* and words following an adverb, F (1, 34) = 0.901, $p = 0.349$, nor for a Predictability x DM interaction effect, F (1, 34) = 0.330, $p = 0.570$. Note, however, that the waveforms suggested considerable component overlap between N400 and post-N400 effects, as observed elsewhere previously (e.g. Hagoort, 2003; Hoeks & Brouwer, 2014; Kutas, Van Petten, & Kluender, 2006). It is possible that the earlier N400 difference between predictable and unpredictable words obscured any later Predictability effects on post-N400 positivities. Indeed, post-hoc analyses taking N400 amplitude differences into account did suggest effects in the expected direction; these are reported in the Appendix.

With respect to the posterior post-N400 positivity, we hypothesised to find increased amplitudes for pragmatically anomalous dialogue continuations, that is, when *eigenlijk* (signalling unexpectedness) was followed by a contextually predictable word. Analyses of mean amplitudes in the post-N400 time window provided no evidence for such a DM by Predictability interaction, F (1, 34) = 0.080, $p = 0.779$. We did find more positive-going waveforms by 0.67 µV (95% CI [0.37, 0.96], $d_z = 0.54$) in response to words following *eigenlijk* compared with words following an adverb, F (1, 34) = 11.747, $p = 0.002$. This DM effect tended to interact with Hemisphere, F (1, 34) = 3.682, $p = 0.063$; pairwise t tests between mean amplitudes revealed a larger effect of *eigenlijk* over the right hemisphere, t (34) = 3.88, $p < 0.001$, than over the left hemisphere, t (34) = 2.48, $p = 0.018$. These results suggest that the effect of *eigenlijk* on the post-N400 posterior positivity elicited by subsequent words may be independent of their predictability.

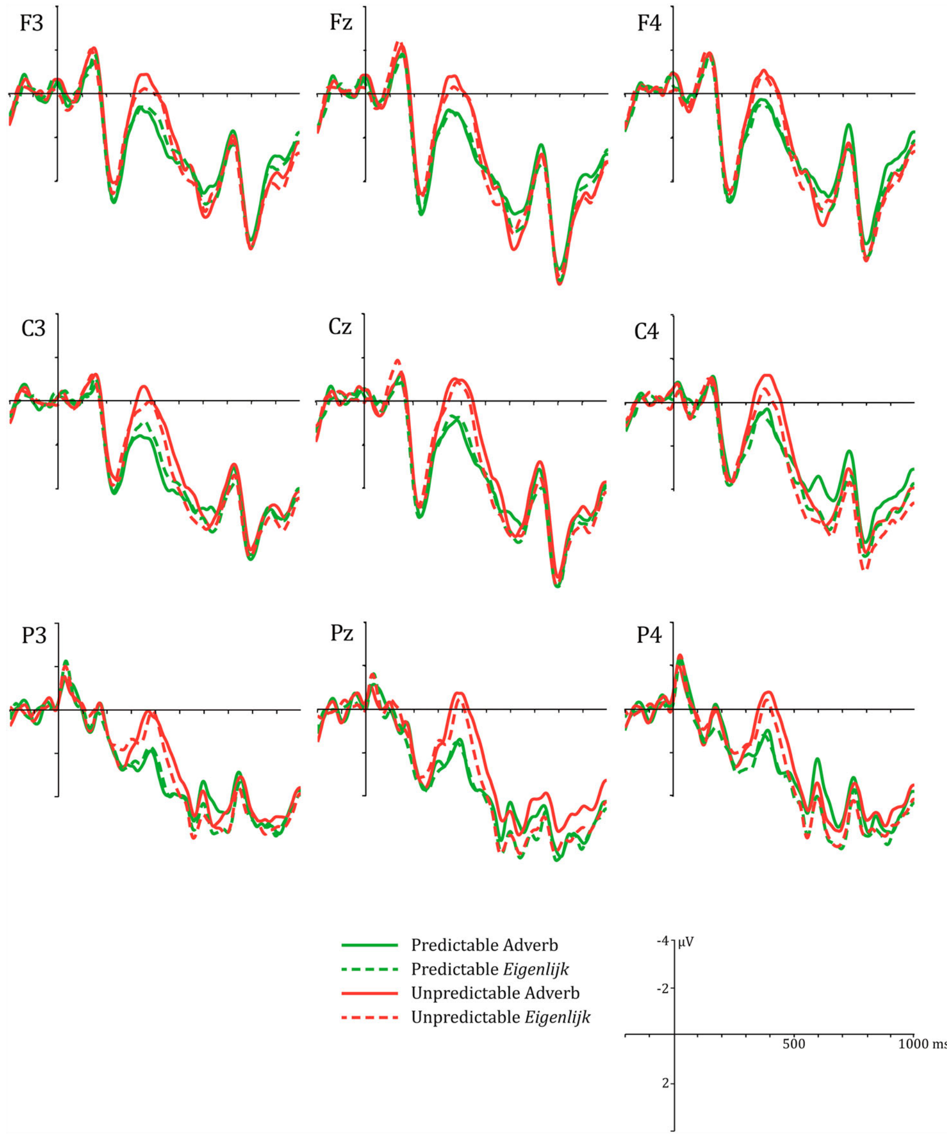

Figure 1. Grand-average ERPs time-locked to critical words at 9 scalp electrode sites (Experiment 1). Critical words were plain-predictable (green lines) or plain-unpredictable (red lines), and followed an adverb (solid lines) or *eigenlijk* (dashed lines). Negative is plotted up in all ERP figures.

Taken together, the ERP findings from Experiment 1 corroborate earlier findings that a word's plain predictability modulates N400 amplitude, but we found no evidence that *eigenlijk* attenuated or reversed this N400 effect. Standard analyses provided no evidence for anterior post-N400 positivity modulations, but revealed that the presence of *eigenlijk* affected incremental processing of subsequent words, such that words following *eigenlijk* elicited an enhanced posterior positivity compared with words following an adverb.

Behavioural memory performance

The percentage of critical words that had been correctly recognised from the reading experiment (73%) was larger by 57% (95% CI [53.0, 61.4], $d_z = 5.78$) than the percentage of false alarms to unseen words (16%). This difference yielded an average d_a of 1.64. The fact that participants readily distinguished between seen and unseen words suggests that they had been paying attention during the reading experiment.

For words that appeared in the experimental dialogues, discriminability ($d_a = 1.75$) was not influenced by Predictability, Discourse Marker, or their interaction (all $F < 0.007$). We hence found no evidence that plain predictability or the presence of *eigenlijk* in the reading experiment affected participants' word recognition memory.

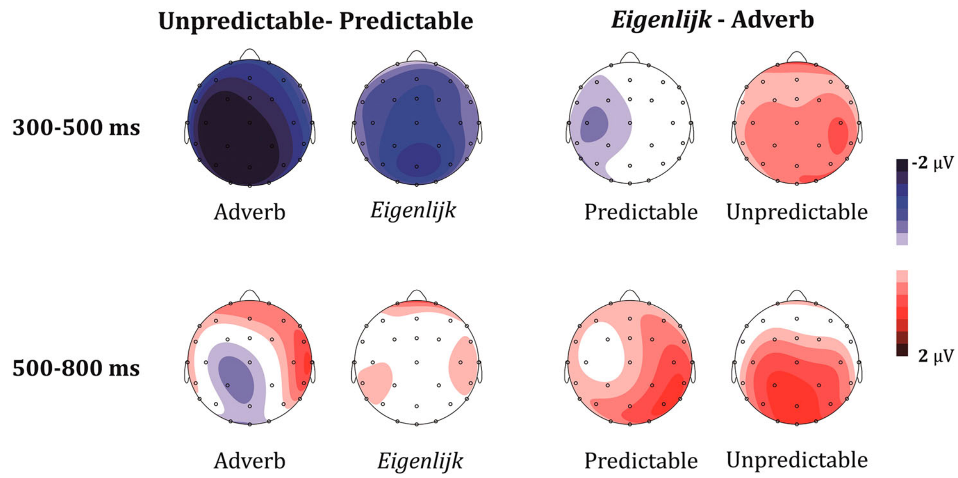

Figure 2. Scalp topographies of the difference in N400 amplitude (upper panel) and in post-N400 positivity (lower panel) in response to plain-predictable vs. plain-unpredictable words (left panels), and in response to words following an adverb vs. *eigenlijk* (right panels).

Experiment 2 (inderdaad)

Event-related potentials

The ERPs time-locked to the presentation of the critical words are presented in Figure 3; Figure 4 shows the scalp topographies of the difference waves between conditions in the N400 and post-N400 positivity time window.

As in Experiment 1 and many previous studies, the N400 elicited by predictable words was smaller compared with the N400 elicited by unpredictable words, by 0.62 µV (95% CI [0.29, 0.97], $d_z = 0.29$), *F* (1,33) =5.914, *p* = 0.021. Contrary to our hypothesis that *inderdaad* would increase this Predictability effect, we found no evidence for a Predictability x Discourse Marker interaction, *F* (1,33) = 0.022, *p* = 0.884. Neither did we find evidence for a difference in N400 amplitude between words following *inderdaad* compared to words following an adverb, F (1,33) = 0.239, *p* = 0.628.

Corroborating evidence from previous studies, the anterior post-N400 positivity was more positive-going by 0.60 µV (95% CI [0.15, 1.18], $d_z = 0.36$) for unpredictable words when compared with predictable words, *F* (1,33) = 4.359, *p* = 0.045. Although most previous studies observed a left frontal maximum (e.g. DeLong et al., 2014), the scalp topography in Figure 4 suggested that the Predictability effect in the current study was right-lateralized. This was confirmed by a Predictability by Hemisphere interaction, *F* (1,33) = 6.320, *p* = 0.017; pairwise t-tests revealed a difference between predictable and unpredictable words over the right hemisphere, *t* (36) = 2.77, *p* = 0.009, but not over the left hemisphere, *t* (36) = 1.18 *p* = 0.245. Our hypothesis that *inderdaad* would modulate this frontal positivity effect was not confirmed: we found no evidence for a Predictability x Discourse Marker interaction, *F* (1,36) = 0.678, *p* = 0.416, nor for a main effect of Discourse Marker, *F* (1,33) = 0.325, *p* = 0.573.

As for the posterior post-N400 positivity, we hypothesised to find increased amplitudes for pragmatically anomalous dialogues, that is, for unpredictable words following the confirmatory cue *inderdaad*. However, results from the posterior post-N400 positivity analysis revealed no evidence for such a Discourse Marker by Predictability interaction, *F* (1,33) = 0.007, *p* = 0.932, nor for a main effect of Discourse Marker, *F* (1,33) = 0.108, *p* = 0.744. We did find more positive-going waveforms by 0.62 µV (95% CI [0.29, 0.96], $d_z = 0.45$) in response to unpredictable words compared with predictable words, *F* (1,33) = 7.371, *p* = 0.001. This Predictability effect was modulated by Hemisphere, *F* (1,33) = 9.920, *p* = 0.003; pairwise t-tests revealed a larger Predictability effect over the right hemisphere, *t* (33) = 3.27, *p* = 0.003, than over the left hemisphere, *t* (33) = 1.98, *p* = 0.056, suggesting a posterior extension of the anterior Predictability effect.

In sum, the ERP findings from Experiment 2 corroborate earlier findings that a word's predictability modulates the N400 and post-N400 anterior positivity, but we found no evidence that the pragmatic confirmation encoded in *inderdaad* enhanced predictability effects on subsequent words online, nor that the presence of *inderdaad* otherwise affected incremental processing of subsequent words when compared with adverbs.

Behavioural memory performance

The percentage of words that had been correctly recognised (73%) was larger by 54% (95% CI [49.9, 58.1], d_z = 5.57) than the percentage of false alarms to unseen words (19%). This difference waspresent in all participants and yielded an average d_a of 1.52.

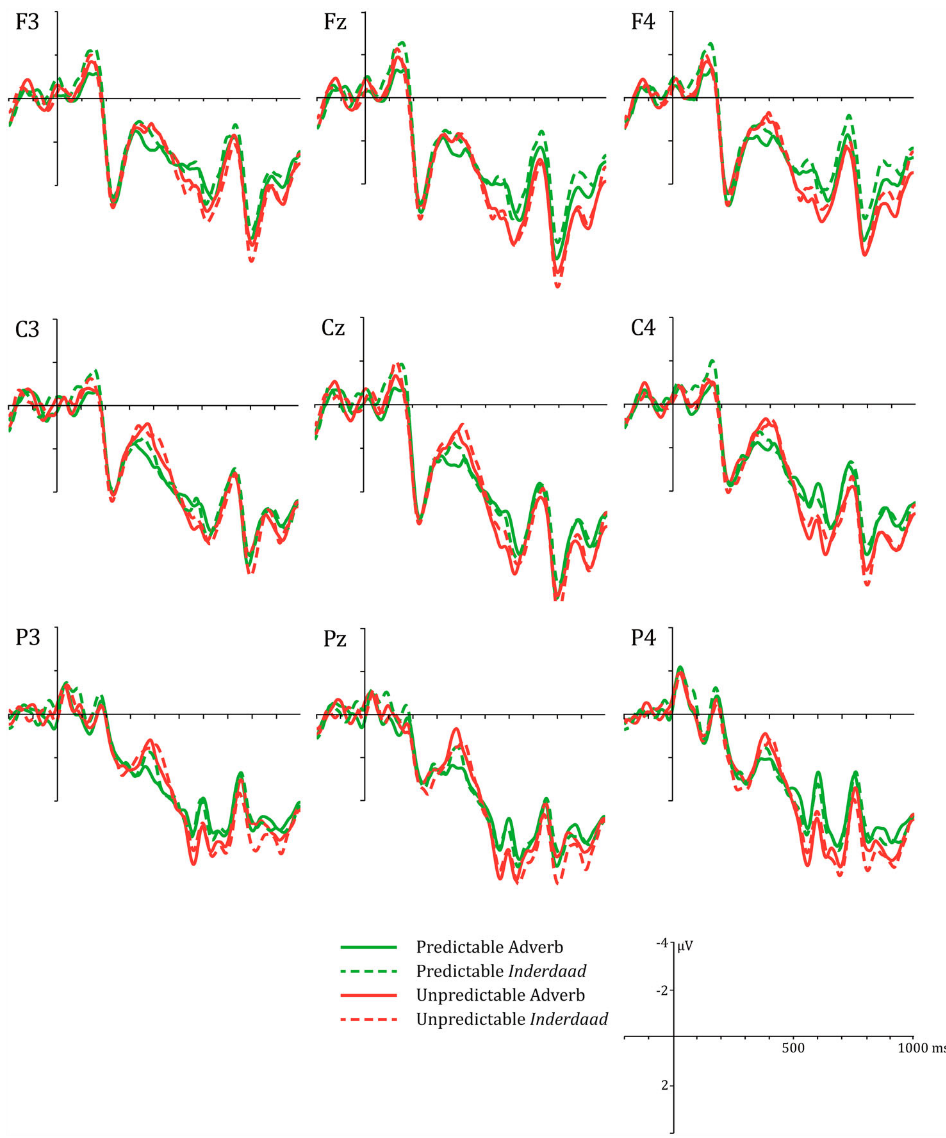

Figure 3. Grand-average ERPs time-locked to critical words at 9 scalp electrode sites (Experiment 2). Critical words were plain-predictable (green lines) or plain-unpredictable (red lines), and followed an adverb (solid lines) or *inderdaad* (dashed lines).

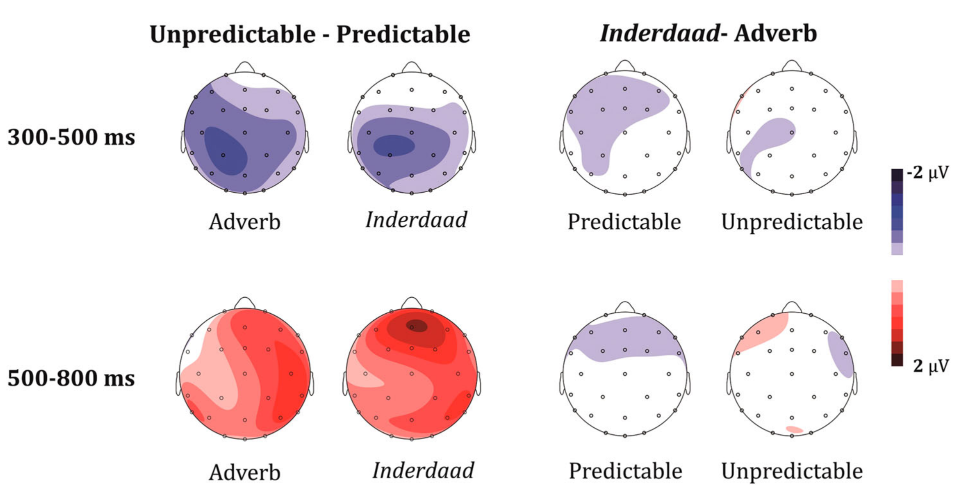

Figure 4. Scalp topographies of the difference in N400 amplitude (upper panel) and in post-N400 positivity (lower panel) in response to plain-predictable vs. plain-unpredictable words (left panels), and in response to words following an adverb vs. *inderdaad* (right panels).

We found no evidence for a difference in memory performance between predictable (1.70) and unpredictable (1.63) words, F (1,39) = 1.551, p = 0.220, but there was an effect of DM indicating better seen/new discriminability for seen words in the presence of *inderdaad* (1.73) than in the presence of an adverb (1.61) (difference 0.12, 95% CI [−0.04, 0.28], d_z = 0.26), F (1,39) = 5.81, p = 0.021. There was no evidence for a Discourse Marker by Predictability interaction effect, F (1, 39) = 0.048, p = 0.828. Thus, results from Experiment 2 suggest that processing words in the presence of *inderdaad* as opposed to an adverb had positive downstream consequences for their accessibility in memory.

Between-experiments comparison

To explore differences between *eigenlijk* and *inderdaad* more directly, we combined the data from both experiments. This revealed that Predictability effects on the N400 in response to critical words were overall attenuated in the *inderdaad*-experiment (0.63 μV) compared with the *eigenlijk*-experiment (1.48 μV); Predictability by Experiment interaction, F (1, 67) = 5.70, p = 0.020 (difference 0.85 μV, 95% CI [0.14, 1.57], d = 0.57). Note that this was even the case for dialogues without a discourse marker, although these dialogues were identical across experiments (difference 1.23 μV, 95% CI [0.13, 1.89], d_s = 0.54), F (1, 67) = 4.95, p = 0.030. This suggests that the regular presence of expectation-managing discourse markers in an experimental context affected semantic processing of subsequent input at a more global, experiment-wide level. There was no Discourse Marker by Experiment interaction or a three-way Predictability by Discourse Marker by Experiment interaction (all F < 0.95).

With respect to later processing stages, combined analyses confirmed a similar lack of evidence for effects of discourse markers on the anterior post-N400 positivity for *inderdaad* and *eigenlijk*, as there were no interactions with Experiment (all F < 1.15). As for the posterior post-N400 positivity, recall that Experiment 1 showed that words following *eigenlijk* elicited a positivity relative to words following a control adverb, and no effects were found for *inderdaad* in Experiment 2. A combined analysis confirmed a Discourse Marker by Experiment interaction, F (1,67) = 4.68, p = 0.034, suggesting that only *eigenlijk* affected incremental processing of subsequent input. There was no evidence for a three-way interaction, F (1,67) = 0.07, p = 0.783.

Finally, regarding recognition memory, a combined analysis of d_a scores across experiments showed a tendency for a Discourse Marker by Experiment interaction, F (1,76) = 3.30, p = 0.073, confirming that only the presence of *inderdaad* affected recognition memory for subsequently processed words. Follow-up independent t-tests suggested that memory for words that had followed a discourse marker did not differ between experiments, t (67) = 0.20, p = 0.839 (difference 0.02, 95% CI [−0.28, 0.35], d_s = 0.03); rather, recognition memory for words occurring in dialogues with an adverb was worse in the *inderdaad*-experiment than in the *eigenlijk*-experiment, t (67) = 1.76, p = 0.080 (difference 0.14, 95% CI [−0.04, 0.60], d_s = 0.28), even though materials in these conditions were identical. This again suggests that the regular presence of expectation-managing discourse markers in the experimental context affected processing more globally. There was no evidence for a three-way interaction, F (1,76) = 0.02, p = 0.900.

Discussion

The present study set out to empirically test and further specify theoretical assumptions about the conversation-managing function of discourse markers by examining their effects on the comprehender. In two experiments, we investigated to what extent the pragmatic information encoded in two expectation-managing discourse markers – *eigenlijk*, marking upcoming prediction disconfirmation, and *inderdaad*, marking upcoming prediction confirmation – affected processing of subsequent (un)predictable input. We hypothesised that, upon encountering a discourse marker (relative to a control adverb in the baseline conditions), comprehenders would adjust their initial expectations about likely dialogue continuations during reading. As such, the presence of *inderdaad* was hypothesised to increase, and the presence of *eigenlijk* to reduce or alter predictability effects, as measured by modulations of N400 and post-N400 positivity amplitudes elicited by subsequently read words.

The results from control dialogues with an adverb corroborated earlier findings that a word's predictability modulates both N400 and post-N400 anterior positivity amplitudes. The topographic distribution of predictability effects on the post-N400 anterior positivity in our study did differ from previous findings: whereas most previous studies reported a left frontal maximum (e.g. DeLong et al., 2014), the predictability effect in Experiment 2 was right-lateralized. We are hesitant to derive strong conclusions from scalp topographical details, but speculatively relate the right-hemispheric bias to the pragmatic predictability manipulation in our experiments: previous studies have reported right-biased late positivities in relation to jokes (Coulson & Kutas, 2001; Coulson & Lovett, 2004) and indirect requests (Coulson & Lovett, 2010), which similarly have a pragmatic basis.

The results provided no evidence for *eigenlijk*- or *inderdaad*-based modulations of predictability effects on the N400, despite the fact that both discourse markers affected predictability in offline cloze probabilities, which are known to be predictive of N400 amplitude (e.g. Kutas & Hillyard, 1984). Apparently, *inderdaad* did not cause comprehenders to strengthen the likely prediction, nor did *eigenlijk* cause comprehenders to discard the likely prediction or change it to the unexpected word to an extent measurable in the N400. To the extent that initial predictions were not discarded, this is consistent with previous studies reporting lingering predictions or interpretations (e.g. Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Corley, 2010; Lowder & Ferreira, 2016; Rommers & Federmeier, 2018b). Another possible explanation for the observed lack of modulation of the N400 by discourse markers is that the unpredictable words were selected from completions of dialogues containing *eigenlijk*. This ensured that the stimuli were relatively natural, but as a side effect the unpredictable words were often semantically related to the predictable words: pairwise semantic similarity (LSA) values obtained on the basis of their English translations (following Chwilla & Kolk, 2002) ranged from 0.05-0.81 (M = 0.25, SD = 0.21). This semantic similarity between predictable and unpredictable words may have reduced N400 amplitude (e.g. Federmeier & Kutas, 1999). For instance, when *eigenlijk* marked the unexpectedness of a colour (e.g. *green*), but the alternative was also a colour (e.g. *grey*), any switch from an original to a revised prediction may have had only minor consequences at the level of the N400. A final possibility is that refining or revising initial predictions may involve computations that require more time than participants had between encountering the discourse marker and reading the critical word (for evidence that contextual facilitation takes time, see Camblin, Ledoux, Boudewyn, Gordon, & Swaab, 2007; Chow, Lau, Wang, & Phillips, 2018; Wlotko & Federmeier, 2015). Future studies could investigate whether discourse markers can modulate N400 predictability effects if more time is available for prediction. This could be done by using slower stimulus presentation or, in an ecologically more valid way, by presenting discourse markers in sentence-initial position (e.g. *eigenlijk hebben we alleen maar regen gehad* "[actually] have we nothing but rain had").

Regarding the post-N400 anterior positivity, we found no evidence for *inderdaad*- or *eigenlijk*-based modulations of predictability effects, hence not supporting the hypotheses that the presence of *inderdaad* increases or the presence of *eigenlijk* reduces processing costs associated with prediction disconfirmations. One possible reason is component overlap with the preceding N400 effect, a possibility explored in the Appendix. The possibility that comprehenders did not use the discourse markers to adjust their expectations about subsequent words seems inconsistent with recent findings from a visual world eye-tracking study (van Bergen & Bosker, 2018), who investigated effects of *inderdaad* and *eigenlijk* on the processing of (un)predictable dialogue continuations by measuring fixations on imagespresented on a screen. Participants read a constraining context sentence (e.g. *Tineke just got back from her holiday on Ibiza, and she is very tanned*), after which they listened to a question (*The weather must have been great there?)* and an incomplete answer that contained either a control adverb or a discourse marker (*We hebben daar / inderdaad / eigenlijk alleen maar … gehad* "we have there / indeed / actually had nothing but … ", where *sun* would be expected). Participants completed the dialogues by clicking on one of four visually presented referents. Relative to a control adverb, encountering *inderdaad* led to increased fixations on likely discourse referents (i.e. *sun*), whereas encountering *eigenlijk* led to increased visual attention to contextually less likely discourse referents (i.e. *rain*), suggesting that participants immediately integrated the pragmatic cues to modulate their predictions about dialogue continuations. However, the visual world paradigm measures potential sentence interpretations using a limited set of visual objects (for discussion, see Dahan & Tanenhaus, 2004; Henderson & Ferreira, 2004; Huettig, Rommers, & Meyer, 2011). The current ERP study did not constrain potential dialogue interpretations visually, which suggests that effects of *eigenlijk* and *inderdaad* on online processing may disappear if more sentence interpretations and continuations are possible.

In interpreting the lack of predictability effects of discourse markers in the present study, we should consider that discourse markers are notoriously polyfunctional: their interpretation depends on the specific characteristics of the discourse (see e.g. Brinton, 1996; Fischer, 2006; Jucker & Ziv, 1998; Maschler & Schiffrin, 2015). *Eigenlijk* and *inderdaad* may for instance mark the assumed (un)expectedness of a name (e.g. *De koningin van Nederland heet eigenlijk Máxima* "the queen of the Netherlands is actually called Máxima"), an event (e.g. *Na het plassen was ik inderdaad mijn handen* "after peeing I indeed wash my hands") or a speech act (e.g. *Hoe heet je eigenlijk?* "what is your name actually?", marking the social unexpectedness of asking this question; van Bergen, van Gijn, Hogeweg, & Lestrade, 2011). Given that in principle, *inderdaad* and *eigenlijk* can signal (un)expectedness of *any* aspect of the discourse, comprehenders perhaps sometimes adjusted a different expectation, or needed more time to work out which likely expectation the speaker intended to modulate when using the discourse marker.

Experiment 1 did reveal an enhanced *posterior* post-N400 positivity in response to words following *eigenlijk*, although this effect was not restricted to plain-predictable and hence pragmatically anomalous words as we hypothesised. An important possibility is that the enhanced posterior positivity elicited by words following *eigenlijk* reflects the processing costs associated with pragmatically driven discourse updating or integration (e.g. Brouwer et al., 2012; Schumacher, 2013). After having encountered a warning signal for upcoming unexpectedness, the comprehender needs to integrate any subsequent input in relation to this adversative cue: after all, any next word may express the crucial unexpected information. It is thus possible that the mere presence of *eigenlijk* required additional processing of *any* dialogue continuation, irrespective of the (pragmatic) predictability of the subsequent information. This proposal would be compatible with the eye-tracking findings reported in van Bergen and Bosker (2018), who found that the presence of *eigenlijk* slowed down responses, regardless of the preferred dialogue completion; it would also corroborate the idea that arriving at a pragmatically more complex discourse interpretation is costly (see also Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014).

Findings from Experiment 2 suggested that, in contrast with *eigenlijk*, the presence of *inderdaad* did not affect online processing of subsequent input. We speculate that this distinction between *eigenlijk* and *inderdaad* reflects a difference in cue informativity: a warning signal for unexpectedness (*eigenlijk*) is arguably a more informative cue to the comprehension system than an advance confirmation of a likely expectation (*inderdaad*). Again considering the polyfunctionality of discourse markers, our findings suggest that *inderdaad* may not typically be used to facilitate rapid on-line processing of subsequent input. Another theoretically recognised function of discourse markers is to manage interpersonal relations: speakers use discourse markers to express acknowledgement of, and attention to, their addressee's social identity or "face" (e.g, Brown & Levinson, 1987; Traugott, 2010). Although this will need further research, we speculate that *inderdaad* more likely serves such a socio-pragmatic goal in conversational interaction: speakers may use *inderdaad* to signal interpersonal agreement, with the aim of establishing social coherence.

Results from the between-experiment comparison suggested that the presence of discourse markers in the experimental context affected processing of dialogues that did *not* contain a discourse marker: for dialogues containing an adverb (which were identical across experiments), comprehenders who had regularly encountered *inderdaad* showed reduced Predictability effects on, and worse memory for, subsequently read words relative to comprehenders who had regularly encountered *eigenlijk*. We speculate that the occurrence of the discourse markers in the experiments modulated the overall utility of prediction. Although the proportion of (plain) predictable and unpredictable dialogue endings was the same across experiments, recall that Plain Predictability directly corresponded with Pragmatic Coherence in the case of *inderdaad*, but not in the case of *eigenlijk*. Consequently, in Experiment 1 (*eigenlijk*) three out of four experimental conditions contained some form of "prediction disconfirmation" (in terms of Plain Predictability, Pragmatic Coherence, or both), whereas in Experiment 2 (*inderdaad*), only two out of four conditions did. As a result, the overall probability of predictive success in Experiment 2 (*inderdaad)* was higher than in Experiment 1 (*eigenlijk).* Comprehenders may have been sensitive to these experiment-wide statistics, and adapted their processing accordingly: a lower likelihood of gaining new information in the *inderdaad*-experiment may have encouraged the comprehension system to operate in a top-down "verification mode", at the expense of thoroughly processing the bottom-up input (e.g. Rommers & Federmeier, 2018a; van Berkum, 2010; for research showing that comprehenders adapt their processing to the statistics of the experimental environment, see, e.g. Bradlow & Bent, 2008; Delaney-Busch, Morgan, Lau, & Kuperberg, 2017; Fine, Jaeger, Farmer, & Qian, 2013; Kaschak & Glenberg, 2004; Norris, McQueen, & Cutler, 2003; but see Harrington Stack, James, & Watson, 2018). Although we believe these findings enhance our understanding of the potential role of discourse markers in online language processing, future research could avoid such effects by manipulating discourse marker type within-subjects; this would also allow for a more direct comparison between *eigenlijk* and *inderdaad*.

Finally, findings from the memory test did not provide further evidence that a word's predictability during processing affected later recognition memory. This is perhaps not surprising, given that in our design, unpredictable words were often semantically related to the predictable words and were likely easy to integrate with the context. Interestingly, however, Experiment 2 did show better memory for words if they followed *inderdaad* than if they followed an adverb during the reading phase; we found no such effect for *eigenlijk*. Although this finding is in need of replication, we speculate that *inderdaad* might have provided positive reinforcement during incremental language processing: despite less thorough processing in the *inderdaad*-experiment overall, the confirmation expressed in *inderdaad* may have encouraged integration of subsequent information with existing knowledge (or *schemata*), which is known to improve later memory for that information (e.g. Brewer & Treyens, 1981; van Kesteren, Ruiter, Fernández, & Henson, 2012).

In sum, the current study makes a novel contribution to the literature on predictive language processing by investigating how the presence of pragmatic expectation-managing cues affects incremental dialogue comprehension. We hypothesised that comprehenders would use expectation-managing discourse markers to adjust their expectations about likely dialogue continuations, which in turn would modulate processing of subsequently read (un)predictable words. However, the findings provided no evidence that discourse markers modulated predictability effects elicited by subsequent words at any processing stage. We attributed this to their polyfunctionality, although the lack of evidence for effects of *eigenlijk* on post-N400 positivities was complicated by possible component overlap. We did find that, unlike *inderdaad*, the presence of *eigenlijk* yielded an increased late positivity, and tentatively linked this to integration of subsequent words. This difference between confirmative and adversative cues was explained in terms of informativity to the comprehension system. The presence of *inderdaad* seemed to have-consequences for memory, perhaps because it encouraged integration of subsequently presented input with previously stored knowledge. In addition, our findings raise the possibility that the presence of pragmatic expectation-managing discourse markers modulated the likelihood of predictive success in an experimental context, which in turn affected comprehenders' overall processing behaviour. Taken together, our findings provide a more nuanced understanding of the theoretically assumed function of expectation-managing discourse markers in conversational interaction.

Notes

1. Example of a weakly constraining filler dialogue (Dutch word order is maintained in the translation of the Answer):

   Context: *Marte heeft kort na haar verjaardag haar vriendin Annemarie op bezoek.*

   Shortly after Marte's birthday, her friend Annemarie pays her a visit.

   Question: Annemarie vraagt: wat voor cadeaus heb je gekregen op je verjaardag?

   Annemarie asks: which gifts did you get on your birthday?

   Answer: *Marte zegt: Ik heb toen/eigenlijk/inderdaad van mijn vriendinnen deze tas gekregen.*

   Marte says: I have then/eigenlijk/indeed from my friends this *bag* received.
2. For the first 17 participants in Experiment 1, data acquisition did not include the electrodes for scalp sites Fp1/2, as these were used for measuring blinks and eye-movements. For these participants the average amplitude was calculated over the 8 remaining electrodes; analyses excluding Fp1 and Fp2 for all participants in both experiments yielded similar results.

Acknowledgments

We thank Maarten van den Heuvel for programming help and Birgit Knudsen and Michel-Pierre Jansen for help with data acquisition.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek: [Grant Number 275-89-022 (GvB),275-89-032 (JR)].

References

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam: John Benjamins.

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. doi:10.1016/j.pragma.2004.05.005

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database. [CD-ROM]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. doi:10.1016/j.cognition.2007.04.005

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*, 207–230. doi:10.1016/0010-0285(81)90008-6

Brinton, L. J. (1996). *Pragmatic markers in english: Grammaticalization and discourse functions*. Berlin: Mouton de Gruyter. doi:10.1515/9783110907582.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*, 1318–1352. doi:10.1111/cogs.12461

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. doi:10.1016/j.brainres.2012.01.055

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44. doi:10.1162/jocn.1993.5.1.34

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Cairns, H. S., Cowart, W., & Jablon, A. D. (1981). Effects of prior context upon the integration of lexical information during sentence processing. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 445–453. doi:10.1016/s0022-5371(81)90551-x

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research*, *1146*, 172–184. doi:10.1016/j.brainres.2006.07.033

Canestrelli, A. R., Mak, W. M., & Sanders, T. J. M. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, *28*(9), 1394–1413. doi:10.1080/01690965.2012.685885

Chow, W.-Y., Lau, E., Wang, S., & Phillips, C. (2018). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience*, *33*(7), 803–828. doi:10.1080/23273798.2018.1427878

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. doi:10.1006/cogp.2001.0752

Chwilla, D. J., & Kolk, H. H. J. (2002). Three-step priming in lexical decision. *Memory & Cognition*, *30*(2), 217–225. doi:10.3758/bf03195282

Clark, H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. doi:10.1016/S0010-0277(02)00017-3

Corley, M. (2010). Making predictions from speech with repairs: Evidence from eye movements. *Language and Cognitive Processes*, *25*(5), 706–727. doi:10.1080/01690960903512489

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. doi:10.1016/j.cognition.2006.10.010

Coulson, S., & Kutas, M. (2001). Getting it: Human event-related brain responses to jokes in good and poor comprehenders. *Neuroscience Letters*, *316*(2), 71–74. doi:10.1016/s0304-3940(01)02387-4

Coulson, S., & Lovett, C. (2004). Handedness, hemispheric asymmetries, and joke comprehension. *Cognitive Brain Research*, *19*(3), 275–288. doi:10.1016/j.cogbrainres.2003.11.015

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.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498–513. doi:10.1037/0278-7393.30.2.498

Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. (2017). Comprehenders rationally adapt semantic predictions to the statistics of the local environment: A Bayesian model of trial-by-trial N400 amplitudes. In E. Davelaar, G. Gunzelmann, H. Howes, & T. Tenbrink (Eds.), *Proceedings of the 39th Annual Conference of the cognitive science society* (pp. 1–6). London: Cognitive Science Society.

DeLong, K., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162. doi:10.1016/j.neuropsychologia.2014.06.016

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. doi:10.1111/j.1469-8986.2011.01199.x

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. doi:10.1111/j.1469-8986.2007.00531.x

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. doi:10.1006/jmla.1999.2660

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. doi:10.1016/j.brainres.2006.06.101

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10), e77661. doi:10.1371/journal.pone.0077661

Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin & Review*, *19*(5), 921–928. doi:10.3758/s13423-012-0279-x

Fischer, K. (ed.). (2006). *Approaches to discourse particles*. Amsterdam: Elsevier.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, *29*(2), 320–326. doi:10.3758/bf03194926

Fox Tree, J. E. (2010). Discourse markers across speakers and settings. *Language and Linguistics Compass*, *4*(5), 269–281. doi:10.1111/j.1749-818x.2010.00195.x

Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language*, *40*(2), 280–295. doi:10.1006/jmla.1998.2613

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, *31*(7), 931–952. doi:10.1016/s0378-2166(98)00101-5

Gerwien, J., & Rudka, M. (in press). Expectation changes over time: How long it takes to process focus imposed by German 'sogar'. In I. Recio, L. Nadal, A. Cruz, & O. Loureda (Eds.), *Methodological approaches to discourse markers (Pragmatics and Beyond series)* (pp. 1–20). Amsterdam: John Benjamins Publishing.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*(6), 883–899.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441. doi:10.1126/science.1095455

Harrington Stack, C. M., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & Cognition*, *46*(6), 864–877. doi:10.3758/s13421-018-0808-6

Henderson, J. M., & Ferreira, F. (2004). Scene Perception for Psycholinguists. In J. M. Henderson, & F. Ferreira (Eds.), *The* *interface of language, vision, and action: Eye movements and the visual world* (pp. 1–58). New York, NY: Psychology Press.

Hoeks, J. C. J., & Brouwer, H. (2014). Electrophysiological research on conversation and discourse processing. In T. M. Holtgraves (Ed.), *The Oxford handbook of language and social cognition* (pp. 365–386). New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780199838639.013.024.

Huettig, F., Rommers, J., & Meyer, A. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. doi:10.1016/j.actpsy.2010.11.003

Jucker, A. H., & Ziv, Y. (eds.). (1998). *Discourse markers: Descriptions and theory*. Amsterdam: John Benjamins.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647–670. doi:10.1111/j.1749-818x.2008.00072.x

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, *133*(3), 450–467. doi:10.1037/0096-3445.133.3.450

Kim, C. S., Gunlogson, C., Tanenhaus, M. K., & Runner, J. T. (2015). Context-driven expectations about focus alternatives. *Cognition*, *139*, 28–49. doi:10.1016/j.cognition.2015.02.009

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. doi:10.1016/j.brainres.2006.12.063

Kuperberg, G., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language And*, *21*(5), 489–530. doi:10.1080/01690960500094279

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. doi:10.1080/23273798.2015.1102299

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, *133*(2), 335–342. doi:10.1016/j.cognition.2014.05.017

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195395518.003.0065.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. doi:10.1016/s1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. doi:10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. doi:10.1126/science.7350657

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. doi:10.1038/307161a0

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In M. J. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd ed., pp. 659–724). London: Academic Press. doi:10.1016/b978-012369374-7/50018-3.

Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies: Evidence from the visual-world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1400–1416. doi:10.1037/xlm0000256

Maschler, Y., & Schiffrin, D. (2015). Discourse markers: Language, meaning, and context. In D. Tannen, H. Hamilton, & D. Schiffrin (Eds.), *The Handbook of discourse analysis* (pp. 189–221). Chichester: John Wiley & Sons. doi:10.1002/9781118584194.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *The American Journal of Psychology*, *63*(2), 176–185. doi:10.2307/1418920

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. doi:10.1006/brln.2001.2588

Mosegaard-Hansen, M.-B. (1998). *The function of discourse particles*. Amsterdam, The Netherlands: John Benjamins. doi:10.1075/pbns.53.

Nieuwland, M. S. (2015). The truth before and after: Brain potentials reveal automatic activation of event knowledge during sentence comprehension. *Journal of Cognitive Neuroscience*, *27*(11), 2215–2228. doi:10.1162/jocn_a_00856

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., … Von Grebmer Zu Wolfsthurn, S. (in press). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*. doi:10.1101/267815

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*(3), 324–346. doi:10.1016/j.jml.2010.06.005

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is too hard to handle. *Psychological Science*, *19*(12), 1213–1218. doi:10.1111/j.1467-9280.2008.02226.x

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. doi:10.1016/j.cognition.2011.09.001

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. doi:10.1162/jocn.2006.18.7.1098

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 201–238. doi:10.1016/S0010-0285(03)00006-9

O'Brien, E. J., & Myers, J. L. (1985). When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 12–21. doi:10.1037/0278-7393.11.1.12

Otten, M., & van Berkum, J. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, *45*(6), 464–496. doi:10.1080/01638530802356463

Perry, A. R., & Wingfield, A. (1994). Contextual encoding by young and elderly adults as revealed by cued and free recall. *Aging, Neuropsychology, and Cognition*, *1*(2), 120–139. doi:10.1080/09289919408251454

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. doi:10.1017/s0140525(12001495

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. doi:10.1162/jocn.2010.21411

Riggs, K. M., Wingfield, A., & Tun, P. A. (1993). Passage difficulty, speech rate, and age differences in memory for spoken text: Speech recall and the complexity hypothesis. *Experimental Aging Research*, *19*(2), 111–128. doi:10.1080/03610739308253926

Rommers, J., & Federmeier, K. D. (2018a). Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects. *Cortex*, *101*, 16–30. doi:10.1016/j.cortex.2017.12.018

Rommers, J., & Federmeier, K. D. (2018b). Lingering expectations: A pseudo-repetition effect for words previously expected but not presented. *NeuroImage*, *183*, 263–272. doi:10.1016/j.neuroimage.2018.08.023

Schegloff, E. A. (2010). Some other 'uh(m)'s. *Discourse Processes*, *47*(2), 130–174. doi:10.1080/01638530903223380

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

Schourup, L. (1999). Discourse markers. *Lingua. International Review of General Linguistics. Revue internationale De Linguistique Generale*, *107*(3-4), 227–265. doi:10.1016/s0024-3841(96)90026-1

Schumacher, P. B. (2013). When combinatorial processing results in reconceptualization: Toward a new approach of compositionality. *Frontiers in Psychology*, *4*, 1–13. doi:10.3389/fpsyg.2013.00677

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), 1–10. doi:10.1371/journal.pone.0066839

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*(3), 382–392. doi:10.1016/j.ijpsycho.2011.12.007

Traugott, E. C. (2010). (Inter)subjectivity and (inter)subjectification: A reassessment. In K. Davidse, L. Vandelanotte, & H. Cuyckens (Eds.), *Subjectification, intersubjectification and grammaticalization* (pp. 29–71). Berlin: Mouton de Gruyter. doi:10.1515/9783110226102.1.29.

van Bergen, G., & Bosker, H. R. (2018). Linguistic expectation management in online discourse processing: An investigation of Dutch inderdaad 'indeed' and eigenlijk 'actually'. *Journal of Memory and Language*, *103*, 191–209. doi:10.1016/j.jml.2018.08.004

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*, 3877–3892. doi:10.1016/j.jml.2018.08.004

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

van Berkum, J. J. A. (2010). The brain is a prediction machine that cares about good and bad – any implications for neuropragmatics? *Italian Journal of Linguistics*, *22*(1), 181–208.

van Berkum, J. J. A., 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*(4), 580–591. doi:10.1162/jocn.2008.20054

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67–82. doi:10.1162/jocn.2008.21170

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. doi:10.1016/j.tins.2012.02.001

Van Petten, C., & Luka, B. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. doi:10.1016/j.ijpsycho.2011.09.015

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20–32. doi:10.1016/j.cortex.2015.03.014

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*(6), 648–672. doi:10.1080/23273798.2014.995679