CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Prediction differs at sentence and discourse level: An event-related potential study

Ruohan Chang[1,2,3] , Xiaohong Yang[2,3,]* and Yufang Yang[2,3,]*

[1]Beijing Sport University, [2]Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences and [3]University of Chinese Academy of Sciences
*Corresponding author: E-mails: yangyf@psych.ac.cn, yangxh@psych.ac.cn

Abstract

This study used event-related potentials (ERPs) to investigate how predicting upcoming words differ when contextual information used to generate the prediction is from the immediately preceding sentence context versus an earlier discourse context. Four-sentence discourses were presented to participants, with the critical words in the last sentences, either predictable or unpredictable based on sentence- or discourse-level contextual information. At the sentence level, the crucial contextual information for prediction was provided by the last sentence, where the critical word was embedded (e.g., *Xiaoyu came to the living room. She made a cup of lemon tea. Then she sat down in a chair. She opened a box/an album to look at the pictures.*), and at the discourse level by the first sentence (e.g., *Xiaoyu took out a box/an album. She made a cup of lemon tea. Then she sat down in a chair. She leisurely looked at the pictures.*). Results showed reduced N400 for predictable words compared to unpredictable counterparts at sentence and discourse levels and also a post-N400 positivity effect of predictability at sentence level. This suggests that both sentence- and discourse-level semantic information help readers predict upcoming words, but supportive sentence context more than discourse context.

Keywords: prediction; N400; post-N400 positivity; sentence; discourse

Introduction

During language comprehension, previous information across various timescales can be used to predict upcoming words (Van Petten & Luka, 2012). How comprehenders generate predictions during language comprehension has been a heated topic in recent years[1] (Ito et al., 2016; Nieuwland et al., 2018; Nieuwland et al., 2019). However, it remains unclear whether and how prediction differs when the contextual information used to generate the prediction is from immediately preceding sentence context (e.g., *Xiaoyu came to the living room. She made a cup of lemon tea. Then she sat down in a chair. She opened a box/an album to look at the pictures.*) or from earlier discourse context (e.g., *Xiaoyu took out a box/an album. She made a*

cup of lemon tea. Then she sat down in a chair. She leisurely looked at the pictures.).
This issue was addressed in the present study.

Prediction effects in language comprehension has been widely examined
(Federmeier, 2007; Kuperberg & Jaeger, 2016). It has been demonstrated that pre-
diction of upcoming words can be generated not only based on local sentence con-
text (Delong et al., 2005; Federmeier et al., 2007) but also on global discourse context
(Otten & Van Berkum, 2007, 2008; Van Berkum et al., 2005). Highly constraining
contexts (e.g., "in order to see the cells he used a . . . ") or mildly constraining con-
texts (e.g., "in order to see the objects he used a . . . ") have usually been designed to
make an upcoming word (e.g., "microscope") predictable or less predictable. In
electroencephalogram (EEG) studies, predictable words were more associated with
reduced N400 (an ERP component) amplitudes in comparison to less predictable or
unpredictable words (Laszlo & Federmeier, 2009; Li et al., 2017). The N400 com-
ponent is a negative-going wave that reaches its peak at about 400 ms after the onset
of the stimulus (Kutas & Hillyard, 1980, 1984). In prediction research, a reduced
N400 amplitude for predictable words has been considered to reflect the benefits
of confirmed prediction for semantic processing (Van Petten & Luka, 2012).
However, it is worth noting that N400 effects of prediction may not solely reflect
predictive processing but could reflect a combination of processes that include both
prediction and integration (Nieuwland et al., 2019).

Although the N400 effect was consistently observed to be sensitive to the predic-
tion of upcoming words in both sentence and discourse contexts, a frontal post-
N400 positivity has been reported to be sensitive to prediction processing mainly
in sentence context (Delong et al., 2011; Federmeier et al., 2007; Freunberger &
Roehm, 2016; Thornhill & Van Petten, 2012; Van Petten & Luka, 2012). For
instance, Thornhill and Van Petten (2012) found that a larger post-N400 positivity
was elicited by unpredictable words than predictable words in sentence context,
indicating the sensitivity of the anterior positivity to prediction of upcoming words.
Despite these findings, the exact function of this anterior post-N400 positivity
remains debated. One perspective is that this frontal positivity might be related
to the costs of the inhibition of predicted words (Federmeier et al., 2007;
Thornhill & Van Petten, 2012). An alternative perspective is that this late positivity
might be linked to the revision of message- and/or discourse-level information fol-
lowing unpredictable information (Brothers et al., 2015; Freunberger & Roehm,
2016). It is worth noting that a more broadly distributed post-N400 positivity effect
has been consistently observed for semantically incongruent words than for seman-
tically congruent words during language processing and that this post-N400 posi-
tivity effect has been considered to reflect integration cost (Brouwer et al., 2012;
Burkhardt, 2007). Moreover, many other researchers have considered that the
post-N400 positivity associated with semantic incongruence might be related to
domain-general processes (e.g., Sassenhagen & Bornkessel-Schlesewsky, 2015;
Sassenhagen et al., 2014).

While frontal late positivity associated with prediction has been frequently
reported in sentence-level studies, it has been seen in only a few discourse-level
studies (e.g., Brothers et al., 2015; Delong et al., 2014). In light of these findings,
it seems that sentence- and discourse-level contexts may affect prediction of upcom-
ing words differently.

Language processing is incremental: predictions of upcoming words can be formed at various representational levels on the basis of the preceding context, which unfolds over time (Altmann & Mirković, 2009). According to the memory-based view of language comprehension, contextual information does not indefinitely remain in a state of full activation in working memory but instead gradually decays as language unfolds (Albrecht & Myers, 1998; Kintsch, 1988; Van Den Broek et al., 2005). Based on this view, it is likely that (more recent) sentence-level information will have a stronger effect on linguistic prediction compared to (less recent) discourse-level information due to declining memory traces in working memory (Myers et al., 2000; O'Brien et al., 1990).

However, there have also been studies showing that global context led to a stronger effect than local context (Boudewyn et al., 2012; Camblin et al., 2007). In light of this, it would be expected that discourse-level information could facilitate the prediction of upcoming words to a larger extent than sentence-level information. In fact, this has been observed in one prior study. In Boudewyn et al. (2015), participants were presented with highly constraining discourses wherein the critical words were predictable or unpredictable based on the global discourse context and were semantically congruent or incongruent with the local sentence context (e.g., "Frank was throwing a birthday party, and he had made the dessert from scratch. After everyone sang, he sliced up some *sweet/healthy* and tasty *cake/veggies* that looked delicious"). A graded N400 was observed at the critical nouns; that is, the N400 was the smallest for the globally predictable and locally congruent words ("sweet and tasty *cake*"), followed by globally predictable but locally incongruent words ("healthy and tasty *cake*"), then by globally unpredictable but locally congruent words ("healthy and tasty *veggies*"), and finally by globally unpredictable and locally incongruent words ("sweet and tasty *veggies*"). Given that a smaller N400 was observed for globally predictable but locally incongruent words than for globally unpredictable but locally congruent words, the results indicated that global information can have a stronger influence than local information.

Boudewyn et al.'s study is valuable because it revealed the effects of the local sentence and global discourse contexts on the prediction of target words and indicated that global discourse context can have a stronger effect on online generated predictions of upcoming words than local context. However, in that study, sentence contexts provided a prototypical or atypical feature of the target noun, while discourse contexts presented both semantic-level information and a story scenario, which can be more important for discourse comprehension than semantic information (Boudewyn et al., 2012; Boudewyn et al., 2013; Camblin et al., 2007). However, discourse contexts (in general) can provide many kinds of information, and some of them do not describe a story scenario, but detailed story information (Ledoux et al., 2006; Otten & Van Berkum, 2007). It is unclear whether this type of information has an influence on online predictive processing of upcoming words at all. Moreover, in Boudewyn et al.'s study, the predictability of the target words was mainly determined by the global discourse context rather than the local sentence context; more specifically, the global discourse context might not make a target word (e.g., "veggies") very predictable even if the local sentence context contained a feature word (e.g., "healthy") (see Nieuwland, 2019 for a discussion of this issue). Therefore, how

sentence- and discourse-level contextual information influences the online prediction of upcoming words needs and deserves to be explored further.

Using ERPs, the present study aimed to examine the effects of sentence and discourse context on the prediction of upcoming words. Four-sentence discourses with the critical words located in the last sentence were used as materials (see Table 1).[2] For the sentence level, the crucial information that allows prediction of critical words was provided by the last sentence of the discourse, while for the discourse level it was provided by the first sentence. The level of constraint of the sentence and discourse conditions was manipulated so that the critical words were either predictable or unpredictable.[3] Also, these manipulated sentence and discourse contexts held similar, specific information. Thus, the only difference between the sentence and discourse conditions was the distance between the critical words and the preceding contexts, which were critical to predict the critical words. With the context and predictability of critical words orthogonally manipulated, four conditions were constructed: sentence/unpredictable (*discourse/unpredictable*), sentence/predictable (*discourse/unpredictable*), discourse/unpredictable (*sentence/unpredictable*), and discourse/predictable (*sentence/unpredictable*). The critical words, the two words preceding the critical words, and the second and third sentences were held constant in a set of discourses.

We hypothesized that the neurocognitive process of predicting upcoming words would be associated with both benefits for confirmed prediction and costs for disconfirmed prediction, which would be reflected in a N400 and a frontal late positivity effect respectively (Van Petten & Luka, 2012). Following previous studies (Delong et al., 2005; Li et al., 2017; Otten & Van Berkum, 2007; Van Berkum et al., 2005), we hypothesized that both sentence-level and discourse-level supportive context could benefit semantic processing of predictable words, which could then lead to a smaller N400 for the predictable words than for the unpredictable words in both the sentence and discourse conditions. Moreover, the unpredictable words might give rise to the inhibition of predicted words (Federmeier et al., 2007; Thornhill & Van Patten, 2012) or the revision of message- and/or discourse-level information (Brothers et al., 2015; Freunberger & Roehm, 2016), which would be reflected in a frontal late positivity effect. If sentence context has a stronger effect on the prediction of target words than discourse context because of the decayed memory trace of the context (McKoon & Ratcliff, 1998; Myers & O'Brien, 1998), then a larger frontal late positivity effect would be observed in the sentence condition than in the discourse condition. Alternatively, if the discourse context has a stronger impact than sentence context (Boudewyn et al., 2015), then a larger frontal late positivity effect would be observed in the discourse condition than in the sentence condition.

Method

Participants

Twenty-four university students (12 males; mean age = 22.5 years; range = 18.0–30.0 years; SD = 3.1 years) participated in this study as paid volunteers. They were all native speakers of Mandarin Chinese, were right-handed, had normal or

Table 1. Samples of the stimuli used in the experiment

sentence/unpredictable
晓雨 来到 客厅 里, 泡了 杯 柠檬茶, 然后 坐在 椅子 上, 打开 盒子 翻看起 那些 照片 来了。 Xiaoyu/came to/the living room/in, made/a cup of/lemon tea, then/sat down/a chair/in, opened/a box/look at the/**pictures**. Xiaoyu came to the living room. She made a cup of lemon tea. Then she sat down in a chair. She opened a box to look at the **pictures**.
sentence/predictable
晓雨 来到 客厅 里, 泡了 杯 柠檬茶, 然后 坐在 椅子 上, 打开 相册 翻看起 那些 照片 来了。 Xiaoyu/came to/the living room/in, made/a cup of/lemon tea, then/sat down/a chair/in, opened/an album/look at the/**pictures**. Xiaoyu came to the living room. She made a cup of lemon tea. Then she sat down in a chair. She opened an album to look at the **pictures**.
discourse/unpredictable
晓雨 拿出了 盒子, 泡了 杯 柠檬茶, 然后 坐在 椅子上, 慢悠悠地 翻看起 那些 照片 来了。 Xiaoyu/took out/a box, made/a cup of/lemon tea, then/sat down/a chair/in, leisurely/looked at/the/**pictures**. Xiaoyu took out a box. She made a cup of lemon tea. Then she sat down in a chair. She leisurely looked at the **pictures**.
discourse/predictable
晓雨 拿出了 相册, 泡了 杯 柠檬茶, 然后 坐在 椅子 上, 慢悠悠地 翻看起 那些 照片 来了。 Xiaoyu/took out/an album, made/a cup of/lemon tea, then/sat down/a chair/in, leisurely/looked at/the/**pictures**. Xiaoyu took out an album. She made a cup of lemon tea. Then she sat down in a chair. She leisurely looked at the **pictures**.

Note: The critical words are all bolded. The words that have been changed are underlined. The discourses were identical for the sentence and discourse conditions, and the second and third sentences were identical across the four conditions. Commas were used to separate the sentences of each discourse; this is often observed in short discourse in Chinese and would look more natural to Chinese readers. These commas, however, were changed into periods in the English translations to conform to the conventions of English.

corrected-to-normal vision, and were without any reported neurological disorders. They all signed the informed written consent form required by the ethics committee of the Institute of Psychology, Chinese Academy of Sciences. The data of two female participants were excluded from the statistical analysis due to excessive artifacts.

Materials

We used 120 sets of discourses. Each discourse consisted of four sentences; the critical words were presented in the last sentence and were the same across all experimental conditions. The first sentence was manipulated as the discourse context, and the content before the critical word in the fourth sentence was manipulated as the sentence context. The critical word was either predictable or unpredictable based on the preceding sentence or discourse context; in other words, for the sentence condition, the contextual information in the fourth sentence was changed to make the

Table 2. Pretest results for experimental materials

Pretest Mean (SD)	Sentence/ unpredictable	Sentence/ predictable	Discourse/ unpredictable	Discourse/ predictable
Cloze probability (%)	9.25 (11.89)	82.58 (15.42)	9.17 (12.54)	80.75 (14.15)
Constraint (%)	34.25 (16.89)	82.58 (15.42)	32.67 (16.99)	80.75 (14.15)
Acceptability	5.30 (0.76)	5.48 (0.60)	5.28 (0.74)	5.60 (0.58)

critical word predictable or unpredictable, while for the discourse condition, the contextual information in the first sentence was changed to make the critical word predicable or unpredictable. The second and third sentences were the same in each set of discourses; thus, there were four conditions in total: sentence/unpredictable (*discourse/unpredictable*), sentence/predictable (*discourse/unpredictable*), discourse/unpredictable (*sentence/unpredictable*), and discourse/predictable (*sentence/unpredictable*). Following a Latin Squares design, the 120 sets of discourses were arranged into four lists, each containing 30 discourses per condition, to ensure that each discourse was presented only once. In addition, 90 incongruent discourses were added as fillers: 20 containing semantic violations in the last sentence, 20 containing semantic violations between the last sentence and the first sentence, and 50 containing semantic violations across various distances within the second and third sentences. Thus, the incongruent information was presented approximately equally among the four sentences. This violation paradigm was only used in the fillers, to draw our participants' attention to each sentence equally, and the participants were not asked to detect the violations.

A group of 40 participants (20 males; mean age = 22.0 years; range = 18.0–34.0 years; SD = 3.6 years) who did not participate in the EEG experiment were recruited to complete a cloze probability pretest, in which the discourses were truncated before the critical words and the participants were asked to complete these discourses according to the preceding context. Each participant completed only one of the four lists. The cloze probabilities of the critical words were calculated as the proportion of participants who used the designated critical words to complete these discourses. A repeated-measures ANOVA was conducted with context (2 levels: sentence vs. discourse) and predictability of the critical words (2 levels: predictable vs. unpredictable) as independent variables and cloze probability as the dependent variable. The results revealed a significant main effect of predictability ($F(1, 119) = 2563$, $p = .000$, $\eta^2_p = .96$), indicating that the cloze probability of the predictable words was higher than that of the unpredictable words (see Table 2). The results did not indicate a significant main effect of context or a significant interaction between context and predictability.

Using the cloze probability test, we computed the degree of constraint of the discourses; in this regard, we calculated the largest proportion of participants who were using the same words (critical words and noncritical words) to complete the discourses. Repeated-measures ANOVA revealed a significant main effect of predictability ($F(1, 119) = 690.83$, $p = .000$, $\eta^2_p = .85$), indicating that the constraint of the

predictable conditions was higher than that of the unpredictable conditions (see Table 2). We did not find a significant main effect of context or a significant interaction between context and predictability.

An additional group of 24 university student participants (12 males; mean age = 21.2 years; range = 18.0–25.0 years; SD = 1.9 years) was recruited to rate the acceptability of the discourses on a seven-point scale (1 = extremely unacceptable, 7 = fully acceptable). All conditions were rated acceptable, although a repeated-measures ANOVA revealed a significant main effect of predictability (F (1, 119) = 17.21, p = .000, η^2_p = .13), indicating that predictable conditions were rated as more acceptable than unpredictable conditions (see Table 2). Again, the results showed that there was not a significant main effect of context. These results indicated that acceptability ratings were independent of context but not predictability. We also did not find a significant interaction between context and predictability.

Procedure

Participants were seated comfortably in front of a computer screen in an electrically shielded, sound-attenuating room. The 210 discourses (120 of which were critical materials) were presented in a pseudorandom order with white text on a black background. Each trial began with a fixation cross for 1000 ms in the center of the screen. This was followed by a presentation of the first three sentences of the discourse, sentence by sentence. Participants were asked to press the space bar when they finished reading the sentences. The last sentence was presented one word at a time, and each word appeared for 400 ms and was followed by an interstimulus interval of 300 ms. Participants were instructed not to move or blink during the presentation of individual words. To ensure that the participants read these discourses attentively, 80 out of 210 discourses were followed by a comprehension question related to the preceding discourse content; the participants were told to press "J" or "F" on the keyboard to answer true or false, respectively, and to press the space bar to continue to the trials. After a short practice session of 11 trials, all materials were presented in six blocks (35 trials per block) of about 10 minutes each, with brief rest periods separating the blocks. The entire experiment lasted about 2.5 hours, including preparation.

EEG recording and preprocessing

The EEG data were recorded using a NeuroScan system with an elastic cap containing 64 Ag/AgCl electrodes mounted according to the standard international 10–20 system. Recordings were completed with a sampling frequency of 500 Hz and an amplifier band pass of 0.05–100 Hz. The left mastoid served as the online reference, and the average of both mastoids served as the offline re-reference. The vertical electrooculogram (EOG) was monitored by electrodes placed above and below the participants' left eyes, and the horizontal EOG was recorded by electrodes placed at the participants' right and left outer canthi. All impedances were kept below 5 KΩ during the experiment.

The preprocessing of the EEG data was completed using NeuroScan 4.3 software (NeuroScan Labs, Houston, Texas). EOG artifacts were automatically corrected,

using the ocular artifact algorithm of the NeuroScan 4.3 software package. The data were filtered off-line with a 0.1–30 Hz band-pass filter (24 dB/octave per slope). Critical epochs ranged from 200 ms before the onset of the critical words to 1000 ms after this onset, with baseline of –200 to 0 ms preceding this onset. An artifact rejection procedure was conducted, with a min/max-criterion of ± 75 µV. An average of 4.9% of all trials were rejected, with the remainder being equally distributed among the four conditions. There were, respectively, 28.73 ± 1.75, 28.41 ± 1.65, 28.50 ± 1.22, and 28.50 ± 2.37 artifact-free trials obtained for the sentence/unpredictable, sentence/predictable, discourse/unpredictable, and discourse/predictable conditions, respectively.

Data analysis

Based on visual inspection and previous research (Delong et al., 2014; Federmeier et al., 2007; Freunberger & Roehm, 2016), the time windows of N400 and late positivity were determined to be 300–500 ms and 500–700 ms, respectively. The average amplitude of the ERPs within each of the selected time windows was computed across all trials per condition for each participant. A four-way repeated-measures ANOVA was conducted with context (2 levels: sentence, discourse), predictability (2 levels: predictable, unpredictable), hemisphere (3 levels: left, medial, right), and region (3 levels: frontal, central, parietal) as within-subject variables. Therefore, nine regions of interests were established, each containing five or six electrodes: left frontal (F3, F5, F7, FC3, FC5, FT7); left central (C3, C5, CP3, CP5, TP7); left parietal (P3, P5, P7, PO5, PO7, CB1); medial frontal (F1, FZ, F2, FC1, FCZ, FC2); medial central (C1, CZ, C2, CP1, CPZ, CP2); medial parietal (P1, PZ, P2, PO3, POZ, PO4); right frontal (F4, F6, F8, FC4, FC6, FT8); right central (C4, C6, CP4, CP6, TP8); and right parietal (P4, P6, P8, PO6, PO8, CB2).

We only reported the significant (including marginally significant, i.e., $p < .06$) effects involving the experimental conditions. All significant interactions were followed by simple effect tests. When Mauchly's test of sphericity was significant, the Greenhouse–Geisser correction was applied to correct the reported p-values, but the degrees of freedom were uncorrected.

Results

Behavioral results

The average accuracy rate for the reading comprehension questions was 96.05% ($SD = 2.48$%), indicating that participants read the discourses carefully. Moreover, the average accuracy rates were comparable across the four conditions (mean ± SD: sentence/unpredictable, 97.77% ± 4.41%; sentence/predictable, 95.98% ± 5.80%; discourse/unpredictable, 93.30% ± 8.80%; discourse/predictable, 95.47% ± 7.78%).

ERP results

Figure 1 displays the grand average waveforms elicited by the critical words in all four conditions at nine representative electrodes (F3/FZ/F4, C3/CZ/C4, P3/PZ/P4).
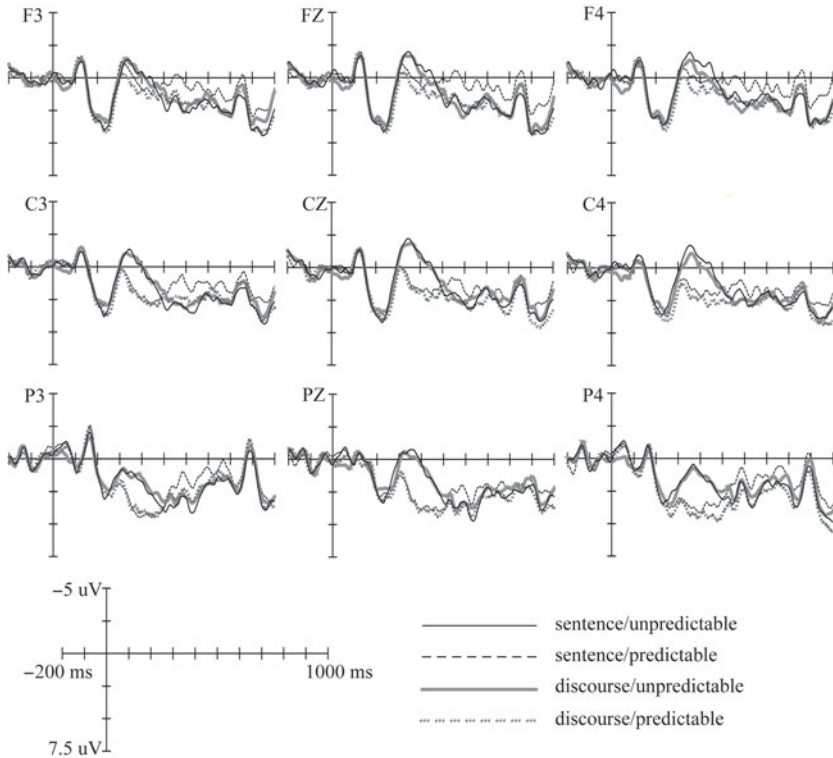
Figure 1. Grand average waveforms evoked by the critical words in the four conditions at nine selected electrode sites. Waveforms are time-locked to the onset of the critical words and negative amplitude is plotted up.

Figure 2 displays the topographies of the difference waves in the N400 (300–500 ms) and late positivity (500–700 ms) time windows. Results of the repeated-measures ANOVAs in the selected time windows are presented in Table 3.

The results of the repeated-measures ANOVAs for the 300–500 ms time window show a significant main effect of predictability (see Table 3), indicating that predictable words reduced N400 amplitude relative to unpredictable words (mean ± SE: predictable, 1.93 ± 0.34 µV; unpredictable: 0.36 ± 0.42 µV; difference [predictable-unpredictable]: 1.57 ± 0.28 µV). There was a significant interaction between predictability and region (see Table 3). A simple effect analysis revealed that the effect of predictability was obtained over all three region levels but most prominently over the central and parietal areas (frontal: F (1, 21) = 6.47, p = .019, η^2_p = .24; central: F (1, 21) = 41.39, p = .000, η^2_p = .66; parietal: F (1, 21) = 86.13, p = .000, η^2_p = .80). There was also a significant interaction between predictability and hemisphere (see Table 3). A simple effect analysis for this interaction revealed that the effect of predictability was obtained over all three hemisphere levels (left: F (1, 21) = 18.25, p = .000, η^2_p = .47; medial: F (1, 21) = 31.85, p = .000, η^2_p = .60; right: F (1, 21) = 33.57, p = .000, η^2_p = .62). There was an interaction between context and region as well (see Table 3); the simple
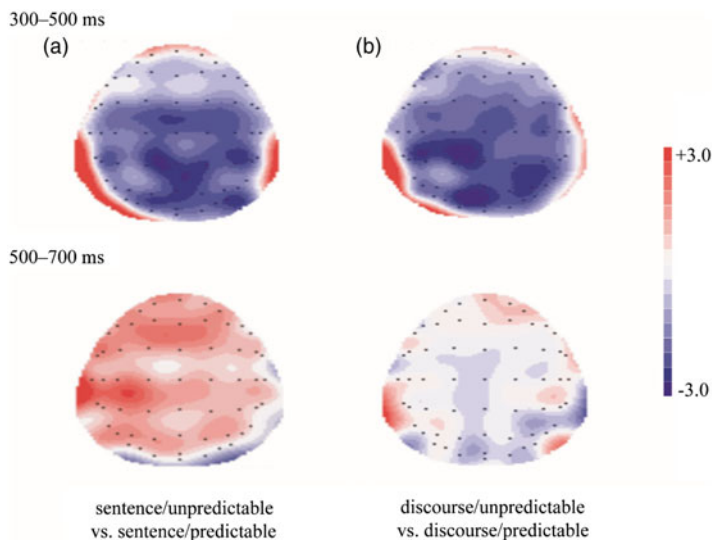
Figure 2. Topographies of the difference waves formed by subtracting ERPs for the sentence/predictable from sentence/unpredictable (a) and discourse/predictable from discourse/unpredictable (b) respectively in the 300-500 ms (top) and 500-700 ms (bottom) time windows.

effect analysis for this interaction revealed that the effect of context was obtained only over the frontal area (F (1, 21) $= 4.77$, $p = .041$, $\eta^2_p = .19$), with the words in the discourse conditions showing reduced N400 amplitude compared to those in the sentence conditions (mean $\pm$ SE: discourse, $0.26 \pm .48$ μV; sentence, -0.24 ± 0.48 μV; difference [discourse-sentence], 0.50 ± 0.23 μV). Moreover, a significant interaction between context and hemisphere was found (see Table 3). The simple effect analysis for this interaction revealed that the effect of context was obtained only over the right hemisphere (F (1, 21) $= 6.29$, $p = .020$, $\eta^2_p = .23$), with the words in the discourse conditions showing reduced N400 amplitudes compared to those in the sentence conditions (mean $\pm$ SE: discourse, 1.32 ± 0.37 μV; sentence, 0.84 ± 0.35 μV; difference [discourse-sentence], 0.48 ± 0.19 μV). No significant interaction between context and predictability was found, indicating that there was no significant difference in the N400 effect between the sentence and discourse conditions.

In the 500–700 ms time window, a significant main effect of context was observed (see Table 3), indicating that the words in discourse conditions elicited more positive-going waves than the words in sentence conditions (mean $\pm$ SE: discourse, 2.13 ± 0.21 μV; sentence, 1.64 ± 0.23 μV; difference [discourse-sentence], 0.49 ± 0.17 μV). An interaction between context and region was obtained (see Table 3). A simple effect analysis revealed that the effect of context was obtained over the frontal and central areas (frontal: F (1, 21) $= 10.99$, $p = .003$, $\eta^2_p = .34$; central: F (1, 21) $= 7.23$, $p = .014$, $\eta^2_p = .26$). More importantly, the results revealed a significant interaction between context and predictability (see Table 3) and a three-way interaction of context $\times$ predictability $\times$ hemisphere (see Table 3). This three-way interaction was broken down by conducting separate analyses for each context

Table 3. ERP results of repeated-measures ANOVAs

Effect	df	300–500 ms			500–700 ms		
		F	p-value	η^2_p	F	p-value	η^2_p
C	1,21	1.63	.216	.07	8.30	**.009**	.28
P	1,21	32.18	**.000**	.61	1.38	.254	.06
C × P	1,21	<1			4.51	**.046**	.18
C × R	2,42	4.90	**.031**	.19	3.88	**.052**	.16
P × R	2,42	9.22	**.005**	.31	1.63	.216	.07
C × P × R	2,42	<1			<1		
C × H	2,42	5.01	**.011**	.19	1.12	.335	.05
P × H	2,42	5.32	**.009**	.20	1.78	.180	.08
C × P × H	2,42	<1			3.33	**.046**	.14
C × R × H	4,84	<1			<1		
P × R × H	4,84	1.58	.208	.07	1.53	.201	.07
C × P × R × H	4,84	<1			1.14	.334	.05

Note: C, context; P, predictability; R, region; H, hemisphere. (Marginally) significant effects are marked in bold.

level. For the sentence level, an ANOVA with predictability and hemisphere as the within-subject factors revealed a significant main effect of predictability (F (1, 21) = 4.76, p = .041, η^2_p = .19) and an interaction between predictability and hemisphere (F (2, 42) = 3.57, p = .051, η^2_p = .15), while the simple effect analysis revealed that unpredictable words elicited more positive-going waves than predictable words over left and medial areas (left: F (1, 21) = 8.69, p = .008, η^2_p = .29; mean ± SE: unpredictable, 2.17 ± 0.30 µV; predictable, 0.85 ± 0.36 µV; difference [unpredictable-predictable], 1.32 ± 0.45 µV; medial: F (1, 21) = 4.53, p = .045, η^2_p = .18; mean ± SE: unpredictable, 2.57 ± 0.42 µV; predictable, 1.53 ± 0.33 µV; difference [unpredictable-predictable], 1.04 ± 0.49 µV). For the discourse level, the ANOVA revealed neither a significant main effect of predictability nor any interaction between predictability and hemisphere (Fs < .37, ps > .695), suggesting that no late positivity effect was present for the discourse condition. Overall, a late positivity effect over the left and medial hemispheres was observed only for the sentence condition, not for the discourse condition.

Discussion

The goal of the present study was to investigate the effects of certain elements of sentence and discourse context on the prediction of target words. To this end, four-sentence discourses were presented in which the last sentence contained a critical word that was either predictable or unpredictable according to the preceding contextual information at sentence or discourse level. ERP results revealed that, relative to the unpredictable words, the predictable words were associated with lower

N400 amplitude in both the sentence and discourse conditions. Also, we found larger late positivity for unpredictable words than for predictable words in the sentence condition but not in the discourse condition. These results indicate that the prediction effect does differ at sentence level from discourse level.

Effect of prediction

In the present study, the predictable words had a reduced N400 in comparison to the unpredictable ones in both the sentence and discourse conditions. This effect was widely distributed across the scalp, most prominently over central and parietal areas, paralleling the scalp distribution of the classic N400 effect (Kutas & Federmeier, 2011). The N400 is considered to reflect the ease or difficulty with which the meaning associated with words can be activated from semantic long-term memory, and a supportive semantic context can facilitate this process (Kutas & Federmeier, 2011). Here, the N400 effect was observed in both the sentence and discourse conditions, suggesting that both contexts can be used to generate a prediction as discourse unfolds, and was found to facilitate semantic processing of predictable words immediately when the prediction was confirmed. It is worth noting that a predictability-dependent N400 is not in itself strong evidence for actual prediction. Recent findings have suggested that predictability-dependent N400 is driven by a cascade of processes that activate and integrate words into contexts (Nieuwland et al., 2019); therefore, the N400 effect in the present study might reflect more than just the prediction of upcoming words.

Interestingly, a late positivity following the N400 also was observed. Unpredictable words elicited a larger late positivity compared to their predictable counterparts in the sentence condition; however, this effect disappeared completely in the discourse condition. This post-N400 positivity in the sentence condition was restrictively distributed over the left and medial electrodes. In previous studies (Delong et al., 2014; Van Petten & Luka, 2012), two kinds of late positive components were observed. Compared to semantically congruent words, semantically incongruent words were associated with a larger positivity with a more posterior distribution (Diza & Swaab, 2007; Pijnacker et al., 2010). However, compared to predictable words, semantically congruent but unpredictable words were also associated with a larger positivity, and this effect was reported to be predominately distributed over more anterior region and was either left-lateralized (Coulson & Van Petten, 2007; Kutas, 1993) or not (Delong et al., 2011; Federmeier et al., 2007). Given that the unpredictable words used in the present study were semantically congruent and a larger late positivity was elicited by the unpredictable words than the predictable words, this late positivity effect might be similar to the anterior positivity found in previous studies (Coulson & Van Petten, 2007; Delong et al., 2011; Federmeier et al., 2007; Kutas, 1993). However, given that our statistical analyses of the late positivity did not show clear anterior distribution, an alternative interpretation is that this late positivity could be the more broadly distributed late positivity, which might reflect a more common integration process between the target words and their preceding context (Brouwer et al., 2012).

Effect of context

The manipulation of context did not show a modulation on the N400 effect in the current study. Based on the N400 results, at least a semantic prediction can be generated in both sentence and discourse contexts, and when the presented information matches the semantic prediction, semantic processing is facilitated comparably in both sentence and discourse conditions. In the sentence condition, the contextual information relevant for predicting the target words is located within the last sentence, which contains the critical word; in the discourse condition, the first sentence of each discourse provided important information for the prediction of the target word, which means that two sentences intervened between the critical word and the contextual information relevant to predict it. Here, the undifferentiated N400 effect between the sentence and discourse conditions may indicate that both contexts could equally facilitate semantic processing of the predictable words.

In the present study, we have carefully constructed our materials so that the cloze probability and acceptability of the critical words were well matched between the sentence and discourse conditions. These two factors have been shown to strongly influence N400 amplitude (Kutas & Federmeier, 2011). The nonsignificant N400 effect between sentence and discourse conditions indicates that these two kinds of contexts have similar impact on semantic processing when the cloze probability and acceptability of words are well matched. Our result is inconsistent with that of Boudewyn and colleagues (2015), who found that the broader discourse context is superior to sentence context for word processing. The discrepancy between the results might be attributed to three factors. First, the sentence context manipulated in Boudewyn et al.'s study provided a prototypical or atypical semantic feature, while the discourse context provided a story scenario, which could have a stronger influence on prediction than lexical semantic information (Boudewyn et al., 2012; Boudewyn et al., 2013; Camblin et al., 2007). In the present study, however, the manipulated contexts for the sentence and discourse conditions contained similar specific lexical information (e.g., box – pictures; album – pictures). Second, in Boudewyn et al. (2015), the cloze probability of the critical words was not matched between the sentence and discourse conditions, whereas it was matched in the present study. Third, spoken stimuli were used as materials in Boudewyn et al. (2015), while written stimuli were used in the present study. Modality could be a significant factor influencing semantic processing (Kutas et al., 1987; Van Petten et al., 1999), which in turn might influence incremental predictive processing (Freunberger & Nieuwland, 2016).

We followed previous studies in using an stimulus onset asynchrony (SOA) of 700 ms to investigate the prediction of upcoming target words (Ito et al., 2016; Li et al., 2017). However, the presentation rate has a significant impact on generation of predictions at different representational levels (e.g., Ito et al., 2016; but see also Delong et al., 2019). The 700 ms SOA might allow more time to generate predictions than in experiments with shorter SOAs, benefiting predictions at more specific representational levels (Ito et al., 2016). In light of this, it is possible that the presentation rate we use may allow our readers to predict more specific information such as lexical items or specific orthographic forms (Pickering & Garrod, 2007, 2013). However, given that the present study was not designed to explore the prediction

of different representational levels, we could not draw a conclusion about whether the lexical items and their specific orthographic form were preactivated. Whether and how presentation rate influences prediction in sentence and discourse context deserves to be explored further in the future.

Interestingly, our results showed that context modulated the late positivity effect. Like the N400, the late positivity was smaller for predictable words than for unpredictable words. However, in contrast to the N400, the late positivity was only found in the sentence condition, not in the discourse condition. This dissociation indicates that the late positivity reflects a different cognitive process than the N400, one that is more sensitive to contextual information.

According to the inhibition view, when the input does not match readers' predictions, the predicted information needs to be inhibited, and the cost associated with this inhibition is reflected in larger frontal late positivity (Kutas, 1993; Van Petten & Luka, 2012). In light of this view, it seems that the larger post-N400 positivity for the unpredictable words than for the predictable words in the sentence condition might reflect the inhibition of predicted words. However, the inhibition account might be untenable given that a larger post-N400 positivity was also observed in the discourse/predictable condition, in which no predictable word needed to be inhibited.

In addition, our results seem to be incompatible with the revision view. According to the revision account, the late positivity is associated with a postlexical, discourse revision mechanism (Brothers et al., 2015; Freunberger & Roehm, 2016). When the encountered input disconfirms the predicted words, the language comprehension system revises the current discourse representation, which leads to an enhanced late positivity in the sentence condition. In light of this, it seems that the observed post-N400 positivity in our study could be related to the revision process. However, it should be noted that in our results, the "outlier" is the sentence/predictable condition. All but the sentence/predictable condition led to enhanced late positivity—including the discourse/predictable condition, in which no unpredictable word initiates the revision process. Therefore, our results seem incompatible with the revision account.

Given that our statistical analysis of the late positivity reveals it to be broadly distributed rather than the frontal post-N400 positivity and also the important result that all but the sentence/predictable condition led to an enhanced late positivity, we consider that the late positivity might reflect the more common integration of upcoming words into preceding context (Brouwer et al., 2012), which might be modulated by context. In the sentence condition, it was more difficult to integrate the unpredictable words into preceding context as compared to the predictable words, which led to a larger late positivity for the unpredictable words than the predictable words. In the discourse condition, however, the relevant context was presented much earlier, and interpreting the target words with respect to this relevant background information have been more cognitively costly, leading to an enhanced late positivity for both the predictable and unpredictable words. In other words, while integrating upcoming words could benefit from detailed semantic information presented late, in the sentence context, this benefit could also be much reduced if the relevant information were presented much earlier, in the discourse context. The reason might be that the memory trace of the discourse context is lower than that of the

sentence context, compatible with the memory-based view (Albrecht & Myers, 1998; Kintsch, 1988; Van Den Broek et al., 2005).

Recently, based on a large-scale (N = 334) study, Nieuwland and colleagues (2019) found that part of the late positivity effect of predictability may be explained in terms of semantic similarity. In addition, there is the possibility that semantic similarity effects are more likely to occur for words that appear close together (here, sentence context, album-pictures) than for distant relationships (discourse context, given that memory trace decreases with longer distance) (Guerra & Knoeferle, 2014). Therefore, we should not exclude the possibility that the positivity effect of predictability that we observed in the sentence condition is at least in part due to a semantic similarity effect. This issue highlights the need to develop a better way to manipulate prediction of target words at sentence and discourse levels in future studies.

There is also an alternative possibility that our results might reflect a priming effect between the critical words and the preceding information in the sentence and discourse context. However, given that a priming effect would not lead to any difference in late positivity in the two sentence conditions (Otten & Van Berkum, 2008), it could be inferred that our results reflect more than a priming effect, that is, also the difference between the prediction of target words at sentence and discourse levels.

In the field of psycholinguistics, difference between sentence and discourse processing has been a central topic for years (Boudewyn et al., 2015; Hasson et al., 2015). Our result adds to the literature by indicating that prediction of upcoming words benefits more from sentence context than from discourse context. This finding is compatible with the findings in previous studies (Myers et al., 2000; Yang et al., 2015; Yang et al., 2018). While those studies did not directly investigate the cognitive process of prediction during language comprehension, they did show that readers need more cognitive effort to integrate long-distance information than short-distance information in discourse context (Myers et al., 2000; Yang et al., 2015; Yang et al., 2018).

Moreover, similarly detailed semantic information was manipulated in the sentence and discourse contexts in the present study to eliminate confounds of contextual priming. It is worth noting that discourse context can provide many kinds of information, including story scenario and detailed semantic information (Ledoux et al., 2006; Otten & Van Berkum, 2007). Our findings add to the literature on prediction in discourse comprehension by indicating that the effect of detailed semantic information might differ from that of a story scenario (e.g., Boudewyn et al., 2015). More importantly, we should be aware that the effect of detailed semantic information might be modulated by the memory trace of contextual information.

In conclusion, the results of this study suggest that both sentence and discourse context allow readers to predict upcoming words and that a supportive sentence and discourse context can facilitate semantic processing of predictable words. More importantly, predictions based on sentence context have a more facilitating effect on processing than predictions based on discourse context. The results indicate that the position of the crucial contextual information that allows the prediction of upcoming words should be taken into consideration when considering prediction during discourse processing.

Notes

1. According to production-based prediction accounts, for instance, language production system can be used to generate prediction of upcoming words and facilitate language comprehension (Pickering & Garrod, 2007, 2013). In addition, comprehenders can predict upcoming words at various linguistic levels, through the preactivation of semantic information, specific form, phonological and orthographic information, etc. A growing number of studies have provided evidence for the production-based prediction account (e.g., Ito et al., 2016; Thornhill & Van Petten, 2012).

2. As shown in Table 1, four-sentence discourses in Mandarin Chinese were used as materials, with commas were used to separate the sentences of each discourse. Although both Chinese and English use periods to indicate sentences' boundaries in text, the Chinese comma functions quite differently from its English counterpart (Kong & Zhou, 2014). In particular, in Chinese, the comma is often used to separate two sentences (Zhou & Xue, 2015); this is more often observed in short stretches of discourse where there is no connecting word or phrase to relate the two sentences (Zhou & Xue, 2012). Thus, using commas to separate the sentences of each discourse in our stimuli would look fairly natural to Chinese readers given that each discourse only contains four sentences.

3. Following previous studies (e.g., Li et al., 2017; Wang et al., 2017), we manipulated contextual constraint to examine predictive processing, under the assumption that it is a good way to tap into predictive processing of target words. However, it should be cautioned that the observed effects of contextual constraint might not purely reflect predictive processing, but could reflect contributions of several factors, including predictability, plausibility, and word similarity (for related discussions, please see Nieuwland et al., 2019).

References

Albrecht, J. E., & Myers, J. L. (1998). Accessing distant text information during reading: Effects of contextual cues. *Discourse Processes*, **26**, 87–107. doi: 10.1080/01638539809545040

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, **33**, 583–609. doi: 10.1111/j.1551-6709.2009.01022.x

Boudewyn, M. A., Gordon, P. C., Long, D., Polse, L., & Swaab, T. Y. (2012). Does discourse congruence influence spoken language comprehension before lexical association? Evidence from event-related potentials. *Language and Cognitive Processes*, **27**, 698–733. doi: 10.1080/01690965.2011.577980

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2013). Effects of working memory span on processing of lexical associations and congruence in spoken discourse. *Frontiers in Psychology*, **4**, 60. doi: 10.3389/fpsyg.2013.00060

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, **15**, 607–624. doi: 10.3758/s13415-015-0340-0

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, **136**, 135–149. doi: 10.1016/j.cognition.2014.10.017

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

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport*, **18**, 1851–1854. doi: 10.1097/WNR.0b013e3282f1a999

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory & Language*, **56**, 103. doi: 10.1016/j.jml.2006.07.005

Coulson, S., & Van Petten, C. (2007). A special role for the right hemisphere in metaphor comprehension? ERP evidence from hemifield presentation. *Brain Research*, **1146**, 128–145. doi: 10.1016/j.brainres.2007.03.008

Delong, K. A., Chan, Wen-hsuan, & Kutas, M. (2019). Similar time courses for word form and meaning preactivation during sentence comprehension. *Psychophysiology*, **56**, e13312. doi: 10.1111/psyp.13312

Delong, K. A., 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**, 1203–1207. doi: 10.1111/j.1469-8986.2011.01199.x

Delong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, **8**, 1117–1121. doi: 10.1038/nn1504

Diaz, M. T., Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*, **1146**, 85–100. doi: 10.1016/j.brainres.2006.07.034

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

Federmeier, K. D., Wlotko, E. W., Ochoa-Dewald, E. D., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, **1146**, 75–84. doi: 10.1016/j.brainres.2006.06.101

Freunberger, D., & Nieuwland, M. S. (2016). Incremental comprehension of spoken quantifier sentences: Evidence from brain potentials. *Brain Research*, **1646**, 475–481. doi: 10.1016/j.brainres.2016.06.035

Freunberger, D., & Roehm, D. (2016). Semantic prediction in language comprehension: Evidence from brain potentials. *Language, Cognition & Neuroscience*, **31**, 1193–1205. doi: 10.1080/23273798.2016.1205202

Guerra, E., & Knoeferle, P. (2014). Spatial distance effects on incremental semantic interpretation of abstract sentences: Evidence from eye tracking. *Cognition*, **133**, 535–552. doi: 10.1016/j.cognition.2014.07.007

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, **19**, 304–313. doi: 10.1016/j.tics.2015.04.006

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory Language & Cognitive Processes*, **86**, 157–171. doi: 10.1016/j.jml.2015.10.007

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, **95**, 163–182. doi: 10.1037/0033-295X.95.2.163

Kong, F., & Zhou, G. (2014). Chinese comma disambiguation on K-best parse trees. In C. Zong et al. (Eds.) *Natural Language Processing and Chinese Computing: Communications in Computer and Information Science 496* (pp. 13–22). Springer-Verlag Berlin Heidelberg.

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

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, **8**, 533–572. doi: 10.1080/01690969308407587

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. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, **207**, 203–205. doi: 10.1126/science.7350657

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

Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing. *Electroencephalography and Clinical Neurophysiology. Supplement*, **39**, 325–330.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory & Language*, **61**, 326–338. doi: 10.1016/j.jml.2009.06.004

Ledoux, K., Camblin, C. C., Swaab, T. Y., & Gordon, P. C. (2006). Reading words in discourse: The modulation of lexical priming effects by message-level context. *Cognitive, Affective, & Behavioral Neuroscience*, **5**, 107. doi: 10.1177/1534582306289573

Li, X., Zhang, Y., Xia, J., & Swaab, T. Y. (2017). Internal mechanisms underlying anticipatory language processing: Evidence from event-related-potentials and neural oscillations. *Neuropsychologia*, **102**, 70–81. doi: 10.1016/j.neuropsychologia.2017.05.017

McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: psycholinguistic research in the 1990s. *Annual Review of Psychology*, **49**, 25–42. doi: 10.1146/annurev.psych.49.1.25

Myers, J. L., Cook, A. E., Kambe, G., Mason, R. A., & O'Brien, E. J. (2000). Semantic and episodic effects on bridging inferences. *Discourse Processes*, **29**, 179–199. doi: 10.1207/S15326950dp2903_1

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, **26**, 131–157. doi: 10.1080/01638539809545042

Nieuwland, M. S. (2019). Do "early" brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience & Biobehavioral Reviews*, **96**, 367–400. doi: 10.1016/j.neubiorev.2018.11.019

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., . . . Von Grebmer Zu Wolfsthurn, S. (2019). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society*. doi: 10.1101/267815

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., . . . Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, **7**, e33468. doi: 10.7554/eLife.33468

O'Brien, E. J., Plewes, P. S., & Albrecht, J. E. (1990). Antecedent retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 241–249. doi: 10.1037/0278-7393.16.2.241

Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, **1153**, 166–177. doi: 10.1016/j.brainres.2007.03.058

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

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, **11**, 105–110. doi: 10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, **36**, 329–347. doi: 10.1017/S0140525X12001495

Pijnacker, J., Geurts, B., Van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, **48**, 2940–2951. doi: 10.1016/j.neuropsychologia.2010.06.003

Sassenhagen, J., & Bornkessel-Schlesewsky, I. (2015). The P600 as a correlate of ventral attention network reorientation. *Cortex*, **66**, A3–A20. doi: 10.1016/j.cortex.2014.12.019

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, **137**, 29–39. doi: 10.1016/j.bandl.2014.07.010

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. doi: 10.1016/j.ijpsycho.2011.12.007

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**, 443–467. doi: 10.1037/0278-7393.31.3.443

Van Den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, **39**, 299–316. doi: 10.1080/0163853X.2005.9651685

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **25**, 394–417. doi: 10.1037/0278-7393.25.2.394

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

Wang, L., Hagoort, P., & Jensen, O. (2017). Language prediction is supported by coupling between frontal gamma and posterior alpha oscillations. *Journal of Cognitive Neuroscience*, **30**, 432–447. doi: 10.1162/jocn_a_01190

Yang, X., Chen, S., Chen, X., & Yang, Y. (2015). How distance affects semantic integration in discourse: Evidence from event-related potentials. *PLoS One*, **10**, e0142967. doi: 10.1371/journal.pone.0142967

Yang, X., Zhang, X., Yang, Y., & Lin, N. (2018). How context features modulate the involvement of the working memory system during discourse comprehension. *Neuropsychologia*, **111**, 36. doi: 10.1016/j.neuropsychologia.2018.01.010

Zhou, Y., & Xue, N. (2012, July). PDTB-style discourse annotation of Chinese text. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers, pp. 69–77). Jeju, Korea.

Zhou, Y., & Xue, N. (2015). The Chinese discourse treebank: A Chinese corpus annotated with discourse relations. *Language Resources and Evaluation*, **49**, 397–431. doi: 10.1007/s10579-014-9290-3