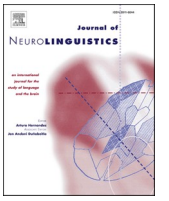
Journal of Neurolinguistics 65 (2023) 101109

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling



Prediction of semantic features is modulated by global prediction reliability: Evidence from the N400 effect

Wenjia Zhang [a,1], Jie Dong [a,1], Xu Duan [a], Yi Zhang [a], Xuefei Gao [b,c,d], Anna Zhen [e], Jie Zhang [f,**], Hao Yan [a,*]

[a] Key Laboratory for Artificial Intelligence and Cognitive Neuroscience of Language, Xi'an International Studies University, Xi'an, China
[b] CAS Key Laboratory of Mental Health, Institute of Psychology, Beijing, 100101, China
[c] Department of Psychology, University of Chinese Academy of Sciences, Beijing, 100049, China
[d] School of Psychology and Counselling, Queensland University of Technology, Brisbane, 4059, Australia
[e] Department of Psychology, St. John's University, 8000 Utopia Parkway, Queens, NY, 11439, USA
[f] Department of Radiation Medicine, Air Force Military Medical University, Xi'an, 710032, China

ARTICLE INFO

Keywords:
N400
Prediction
Semantic similarity
Global prediction reliability
Semantic features

ABSTRACT

Lexical predictability has been shown to be modulated by the global context, but it is unclear whether the global context has a similar modulating effect on the prediction of semantic features. Event-related potentials (ERP) should be helpful in addressing this question, as the N400 effect is sensitive to both lexical predictability and the prediction of semantic features during sentence comprehension. The present study manipulated the semantic similarity between unpredictable target words and expected words in experimental sentences. Different types of filler sentences (predictable vs. incongruous) were used to manipulate the global prediction reliability (GPR). The ERP results showed that, in the predictable filler block, the N400 was reduced when there was high semantic similarity between the unpredictable target word and expected word in the experimental sentences. However, this association was absent in the incongruous filler block, in which participants were discouraged from predicting the upcoming information. These results suggest that the prediction of semantic features of upcoming words could be modulated by the global context.

1. Introduction

There has been a wealth of research on the importance of prediction for human perception, cognition, and action (Clark, 2012). In the field of language comprehension, there is accumulating evidence that readers can generate predictions about upcoming language input. That is, readers can use local contextual (e.g., sentential) information to predict a specific word before its occurrence (Kamide, 2008; Kuperberg & Jaeger, 2016). Moreover, recent studies have revealed that this lexical predictability could be modulated by the global context (Brothers, Dave, Hoversten, Traxler, & Swaab, 2019). The goal of our study was to examine whether the prediction of semantic features of upcoming words could also be modulated by the global context.

* Corresponding author.
** Corresponding author.
E-mail addresses: zhangjie78@fmmu.edu.cn (J. Zhang), haoyan@xisu.edu.cn (H. Yan).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jneuroling.2022.101109
Received 6 August 2021; Received in revised form 6 September 2022; Accepted 11 September 2022
Available online 14 October 2022

Several studies have used event-related potentials (ERP), which have the advantage of high temporal resolution, to examine the prediction mechanism underlying sentence comprehension. In these studies, predictability (cloze probability) of a word and contextual constraint (e.g., sentential constraint) were mainly manipulated. A word's cloze probability is typically empirically defined as the proportion of participants using that word to complete a particular sentence (Taylor, 1953). The sentential constraint is typically empirically defined as the cloze probability of the highest probability continuation (Kutas & Federmeier, 2010). Previous ERP studies have shown that the N400 effect is inversely correlated with a word's cloze probability but less affected by sentential constraint (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Thornhill & Van Petten, 2012; Wlotko & Federmeier, 2012). Because the N400 effect often indexes facilitated lexical-semantic access for more predictable words (Federmeier, McLennan, Ochoa, & Kutas, 2002), the results of these studies suggest that the N400 effect is sensitive to lexical predictability and readers can use sentential context to predict a specific word before its occurrence.

However, it is worth noting that the relationship between N400 and lexical predictability may also be interpreted by the integration view. This is because higher predictability of a specific word intrinsically signifies easier integration of the word into the context. To sum up, the integration view proposes that the N400 effect results from the integration of a word with its previous context (Hagoort, 2008). Alternatively, the lexical view proposes that the N400 effect reflects facilitated lexical access (Lau, Almeida, Hines, & Poeppel, 2009; Lau, Phillips, & Poeppel, 2008). The present study is not aimed to address this debate about the functional interpretations of the N400 effect. Instead, we would adopt the lexical view of the N400 effect to examine the prediction mechanism underlying sentence comprehension.

In addition to predicting a specific word, readers can predict the semantic features of a word, and this semantic prediction also modulates the N400 effect (Federmeier & Kutas, 1999; Wang et al., 2020). For example, Federmeier and Kutas (1999) manipulated three types of critical words: expected word, within-category violation, and between-category violation. Consider the following example: *The gardener really impressed his wife on Valentine's Day. To surprise her, he had secretly grown some* ***roses/tulips/palms.*** "*Roses*" is an expected word, "*tulips*" is a within-category violation, and "*palms*" is a between-category violation. The within-category violation is from the same category as the expected word, whereas the between-category violation is from a different category as the expected word. However, the between-category violation is still a member of a shared higher-level category (i.e., "plants"). It is worth noting that compared with the between-category violation, the within-category violation word shared more semantic features with the expected word, whereas the cloze probability was matched between these two violations. The ERP results showed that the N400 was reduced in the within-category violation compared with the between-category violation. This suggested that pre-activated semantic features of the expected word could facilitate lexical-semantic access for the eventually encountered word.

In real-world communication, sentences are often embedded in global communication contexts with varying statistical properties (Dave, Brothers, Hoversten, Traxler, & Swaab, 2021). Therefore, recent studies aimed to examine whether the strength of top-down predictions could be regulated based on global contextual information (Brothers et al., 2019; Zhang, Chow, Liang, & Wang, 2019). For example, Dave et al. (2021) recruited both young and older adults and asked them to listen to sentences for comprehension. The predictability of a target word was manipulated (highly predictable vs. unpredictable but plausible) in experimental sentences. Global prediction reliability (GPR) was manipulated by adding filler sentences, which ended with words that were either highly predictable or were unpredictable but plausible. The results showed that the N400 effect was sensitive to lexical predictability among both older and younger adults and it was modulated by GPR only among young adults.

Altogether, previous studies have shown that semantic features of a predicted word could be pre-activated (Federmeier & Kutas, 1999; Wang et al., 2020). Moreover, the prediction of a word (sensitivity to lexical predictability) could be modulated by the GPR (Brothers et al., 2019; Dave et al., 2021). Therefore, one would expect that the GPR could modulate the prediction of the semantic features of a target word. However, to our knowledge, no existing studies have empirically examined this gap in the literature, which is the focus of the present study.

1.1. The present study

We manipulated the semantic similarity between the unpredictable target word and the expected word (highest cloze probability) in experimental sentences in the present study. The prediction strength of expected words in experimental sentences was manipulated by using two different types of filler sentences (predictable vs. incongruous). Specifically, the target word in experimental sentences was always unpredictable, but it was semantically congruous to the sentential context. Participants rated the semantic similarity of the unpredictable target word and the expected word in experimental sentences (see Materials for details). The sentences were distributed evenly into two kinds of stimulus blocks, with 50% experimental sentences and 50% filler sentences in each block. In predictable filler blocks, the filler sentences were highly predictable and experimental sentences were unpredictable. Therefore, the GPR in predictable filler blocks was 50% (i.e., 50% of the sentences were predictable). In contrast, in incongruous filler blocks, the filler sentences were semantically incongruous and experimental sentences were still unpredictable. Therefore, the GPR in incongruous filler blocks was 0% (i.e., no sentences were predictable).

We calculated the item-level N400 amplitudes for each experimental sentence across participants. Both analyses of variance (ANOVA) and correlation analyses were used to examine the relationship between the item-level N400 and semantic similarity[2] for each filler type. We first proposed that the N400 amplitude would be reduced with increasing semantic similarity (i.e., an inverse

[2] For simplicity, we define 'semantic similarity' as the semantic similarity between the unpredictable target word and expected word in experimental sentences.

relationship) without global context (Federmeier & Kutas, 1999). More importantly, participants would be more likely to predict the upcoming word as the percentage in GPR increased. Therefore, we predicted that the inverse relationship between the N400 and semantic similarity would be stronger with increasing GPR, such as the inverse relationship would be significant in the predictable filler block (GPR = 50%), but absent in the incongruous filler block (GPR = 0%) (see Fig. 1). This would suggest that the activation of the semantic features of a target word was modulated by the global context.

2. Material and method

Much of the information in this section was also reported in Zhang et al. (2019), Experiment 2.

2.1. Participants

Eighteen participants (6 males, age range: 17–26) from South China Normal University participated in the present study. All participants were right-handed native Mandarin Chinese speakers with normal or corrected-to-normal vision. No participants reported a history of neurological disorders or reading disabilities. All materials and protocols were approved by the Psychology Research Ethics Committee of South China Normal University. Written informed consent was obtained before the experiment. All participants were given a small monetary reward at the end of the study to compensate for their time.

2.2. Materials

A set of 160 Chinese sentences was created, consisting of 80 experimental sentences and 80 filler sentences (see Table 1). The sentence-final two-character words were manipulated to create different kinds of sentences. The experimental sentences were unpredictable but semantically congruous. Forty filler sentences were highly predictable and 40 filler sentences were semantically incongruous.

In the cloze probability rating of sentences, 30 native Mandarin Chinese speakers were asked to provide the most likely word that would continue the sentence for each sentence frame. For the unpredictable experimental sentences, the target words had zero cloze probability. For the predictable filler sentences, the critical words had the highest cloze probability (0.80, SD = 0.09). For the incongruous filler sentences, the critical words were semantically incongruous and always had zero cloze probability. Furthermore, we defined contextual constraint as the cloze probability of the most likely word to complete a given sentence frame. All experimental and filler sentences had a context constraint greater than 0.63. The mean context constraint for experimental sentences of high semantic similarity, experimental sentences of low semantic similarity, highly predictable filler sentences, and semantically incongruous filler sentences were 0.83 (SD = 0.10), 0.81 (SD = 0.11), 0.80 (SD = 0.09) and 0.80 (SD = 0.09), respectively. No significant difference was found across the four types of sentences, F (3, 156) < 1, p > 0.5.

More importantly, 20 participants were recruited to rate the semantic similarity (such as semantic feature overlap) between the unpredictable target word and the expected word (highest cloze probability) in experimental sentences. For example, consider the high SS experiment sentence shown in Table 1. Participants rated the semantic similarity between the unpredictable target word (*牡丹* [*peonies*]) and the expected word (*玫瑰*[*roses*]). The rating was on a 7-point scale (ranging from 1 = minimal similarity to 7 = very strong similarity). The rating scores were first averaged across participants. Based on the cut-off score (3.9), the 80 experimental sentences were divided into the high semantic similarity group such as*牡丹*[*peonies*] - *玫瑰*[*roses*] (>3.9) and the low semantic similarity group such as*钞票*[*cash*] - *星星*[*stars*] (<3.9), with 40 experimental sentences in each group (see Appendix). The average rating score in the high and low semantic similarity groups were 5.05 (SD = 0.78) and 2.42 (SD = 0.84), respectively. The t-test showed that semantic similarity was significantly higher in the high semantic similarity group than the low semantic similarity group, t (78) = 14.54, p < 0.001, CI_{95} = [−2.99; −2.27], d = 0.85.

An additional 20 participants were recruited to rate the semantic plausibility of all sentences. The rating was on a 7-point scale (ranging from 1 = not acceptable at all to 7 = fully acceptable). The average rating score for the experimental sentences of high semantic similarity, experimental sentences of low semantic similarity, predictable filler sentences, and incongruous filler sentences were 5.70 (SD = 0.59), 5.66 (SD = 0.54), 6.61 (SD = 0.45) and 1.67 (SD = 0.59) respectively. The results of ANOVA and further multiple comparison analyses showed that the semantic plausibility of predictable filler sentences was higher than the other three types of sentences, ps < 0.001; the semantic plausibility of both groups of experimental sentences was higher than the incongruous filler sentences, ps < 0.001. More importantly, no significant difference was found between the high semantic similarity and low semantic similarity groups for experimental sentences, t (78) = 0.21, p > 0.8.

The sentences were distributed evenly into four stimulus blocks (two predictable filler blocks and two incongruous filler blocks), with 40 sentences in each block. In a predictable filler block, 20 experimental sentences were intermixed and presented together with 20 predictable filler sentences. In an incongruous filler block, 20 experimental sentences were intermixed and presented together with 20 incongruous filler sentences. Therefore, 50% of the sentences ended with a predictable (high cloze probability) word in the predictable filler blocks (GPR = 50%), and no sentences ended with a predictable word in the incongruous filler blocks (GPR = 0%).

The same type of stimulus blocks was always presented together, and we counterbalanced the order of the type of stimulus blocks across participants such that half of the participants saw the predictable filler blocks first while the other half saw the incongruous filler blocks first. Furthermore, the assignment of the experimental sentences into blocks was also counterbalanced such that a given experimental sentence was presented in a predictable filler block to half of the participants and presented in an incongruous filler block to the other half.

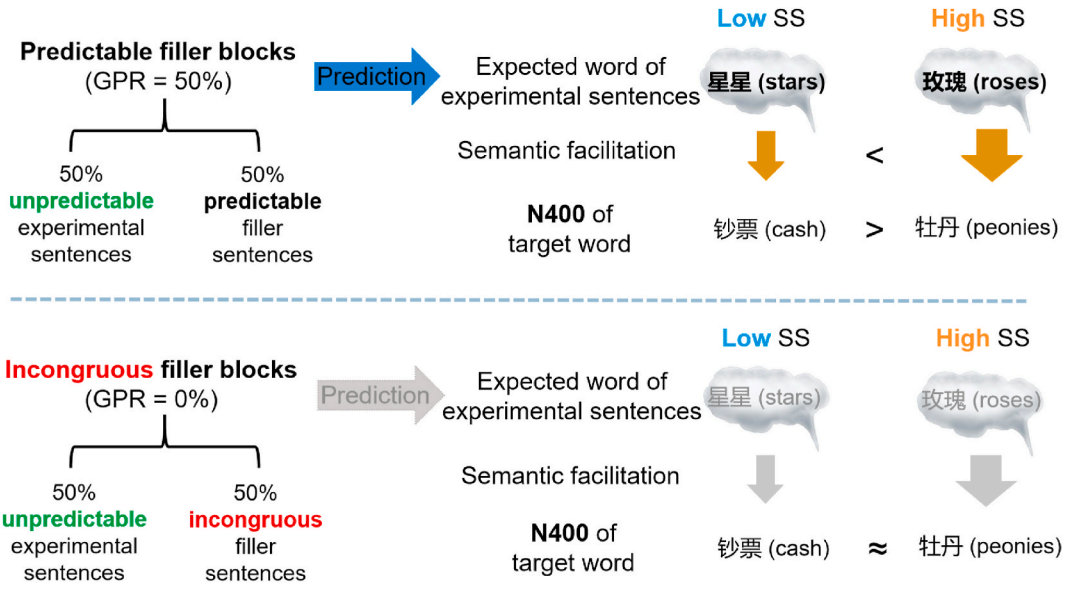


Fig. 1. Prediction of N400 effect at 50% and 0% levels of GPR. Participants were encouraged to predict the expected word in experimental sentences in the predictable filler block (GPR = 50%). The pre-activated semantic features of expected words would facilitate the semantic access of unpredictable target words (and induce a reduced N400) for high SS relative to low SS. However, the proposed mechanism underlying processing would be discontinued at the beginning of the incongruous filler block (GPR = 0%), because participants were discouraged from predicting the expected word in this condition (grey arrows). GPR = global prediction reliability; SS = semantic similarity.

Table 1
Sample experimental and filler sentences.

Sentence Type		Sample Sentence
Experimental sentences	Unpredictable (high SS)	花 店 / 在 情 人 节 /那天/卖出了/ 好多 / **牡丹 (玫瑰)** flower shop/on Valentine's/day/sold/lots of/**peonies (roses)** The flower shop sold lots of **peonies (roses)** on Valentine's Day.
	Unpredictable (low SS)	晚 上 /小孩/ 在 院 子 里 / 数 / **钞票 (星星)** at night/child/in the courtyard/count/**cash (stars)** The child counts the **cash (stars)** in the courtyard at night.
Filler sentences	Predictable	那 辆 / 公 交 车 / 缓 缓 / 驶 离 了 / **车站** the/bus/slowly/left/**station** The bus slowly left the **station**.
	Incongruous	暴风雨 / 前 / 天空 / 布 满 了 / **书包** storm/before/sky/spreads with/**schoolbags** The sky spreads with **schoolbags** before the storm.

Note. The slashes indicate the corresponding segments between Chinese sentences and literal English translations. The words in parentheses are the expected words for sentence frames in experimental sentences. The target words in experimental sentences and the critical words in filler sentences are in bold. Semantic similarity is represented as SS.

2.3. Procedure

The software package E-Prime (Psychology Software Tools, Pittsburgh, PA) was used for stimulus presentation and response collection. All sentences were presented in black font against a light grey background. Each sentence began with a fixation cross for 400 ms and a blank screen for 400 ms. Subsequently, the sentences were presented with word-like units. Each unit appeared on the screen for 400 ms with an inter-stimuli interval (ISI) of 200 ms. After a blank screen of 1200 ms, participants were prompted to rate the semantic plausibility of each sentence on a 7-point scale (ranging from 1 = extremely unacceptable to 7 = fully acceptable). Participants were told to maintain focus on the tasks and avoid eye movements and blinks during the presentation of sentences.

2.4. EEG recording and analyses

Participants were tested individually in a sound-attenuating, electrically shielded booth. Brain Products® system was used for the electroencephalograph (EEG) and electrooculogram (EOG) recording. We used the toolboxes of EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopezcalderon & Luck, 2014) and customized Matlab functions for the EEG analyses. EEG was recorded with 30 Ag/AgCl scalp electrodes (10–20 System) (see Fig. 2) and EOG was recorded from below and above the left eye and at the outer canthus of each eye. The AFz electrode on the cap served as ground. Impedance was kept below 5 kΩ for all electrodes. EEG signals were referenced online to the left mastoid and re-referenced offline to the average of the two mastoid electrodes. The EEG and EOG signals were sampled at 1000 Hz and filtered digitally with a 0.02–30 Hz band-pass offline. Epochs were computed for the 1000 ms after the onset of target words relative to a 200 ms pre-stimulus baseline. Epochs with ocular and movement artifacts (>±80 μV) were rejected.

We first calculated the rejection rate of trials in each type of sentence. The reject rate in predictable fillers, unpredictable targets, and incongruous fillers were 16% (SD = 5%), 16% (SD = 5%) and 15% (SD = 4%) respectively. Repeated measures ANOVA revealed no significant effects for the type of sentence, F (2, 34) = 0.33, p > 0.7. Moreover, we calculated the rejection rate of trials in high and low semantic similarity groups of experimental sentences. The rejection rate in high and low semantic similarity groups were 15% (SD = 10%) and 17% (SD = 10%) respectively. No significant difference was found between high and low semantic similarity groups, t (78) = 0.76, p > 0.4.

To confirm the N400 effect in the present study, we first presented the grand average ERPs elicited by sentence-final target and critical words in three types of sentences: predictable fillers, incongruous fillers, and unpredictable targets (i. e., experimental sentences). Because we had a hypothesis about the locus of the N400 effect (Li, Niefind, Wang, Sommer, & Dimigen, 2015), the N400 amplitude at 300–500 ms was extracted and averaged within an area of interest (AOI) that contained six electrodes (CP3, CPz, CP4, P3, Pz, P4). This AOI covered the central-posterior regions, in which the classical N400 effect was most significant (Kutas & Federmeier, 2010). Repeated measures ANOVA was used to examine the difference of N400 across the three types of sentences.

More importantly, we extracted the item-level N400 amplitude for each unpredictable experimental sentence because we were mainly concerned with how the item-level N400 was modulated by semantic similarity. The N400 amplitudes within the central-posterior AOI and across participants were averaged before statistical analyses. Both ANOVA and correlation analyses were used to examine the relationship between the item-level N400 and semantic similarity, for two levels of GPR. On one hand, we ran a two-by-two ANOVA with GPR levels (50% and 0%) and semantic similarity (high and low) as two factors. On the other hand, the item-level N400 amplitudes were first normalized with z–scores.[3] And then, correlation analyses were used to examine the potential correlation between the N400 and semantic similarity at each level of GPR.

3. Results

3.1. Behavioral results

We first reported the mean scores of semantic plausibility rating of high and low semantic similarity at each GPR level. At the 50% GPR level (predictable filler block), the scores of high and low semantic similarity were 4.48 (SD = 0.70) and 4.41 (SD = 0.80) respectively, no significant difference was found between high and low semantic similarity, t (17) = 0.40, p > 0.6. At the 0% GPR level (incongruous filler block), the scores of high and low semantic similarity were 4.81 (SD = 0.66) and 4.65 (SD = 1.11) respectively, no significant difference was found between high and low semantic similarity, t (17) = 0.65, p > 0.5. Moreover, we reported the mean scores of semantic plausibility rating of all experimental sentences at each GPR level. The scores at the 50% and 0% GPR level were 4.45 (SD = 0.66) and 4.73 (SD = 0.75) respectively, no significant difference was found between the two GPR levels, t (17) = 1.61, p > 0.1.

3.2. ERP results

As shown in Fig. 3, there was a clear negative deflection at 300–500 ms. With comprehensive consideration of its polarity, latency, and topographic distribution, it was confirmed as the N400 component (see supplementary materials for waveforms of all scalp electrodes). The statistical results showed a clear graded N400 effect: incongruous fillers > unpredictable targets > predictable fillers

[3] The normalization of item-level N400 amplitude was just for illustration. By doing this, the scatter points at both GPR levels could be presented within a small range on the Y axis.

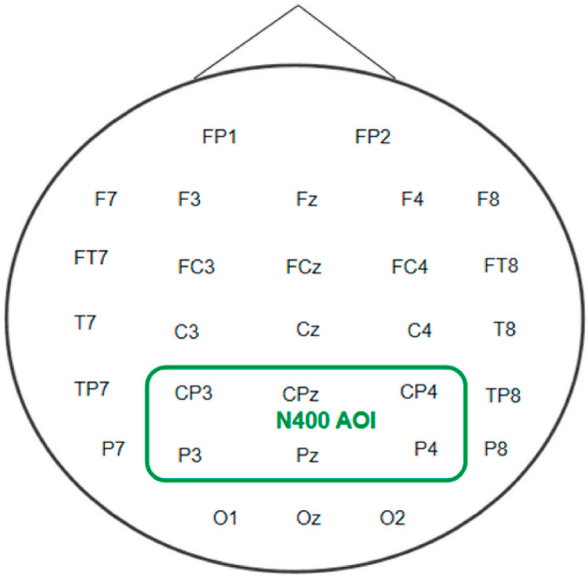


Fig. 2. Thirty recorded scalp electrodes and the AOI for statistical analyses of N400. AOI = area of interest.

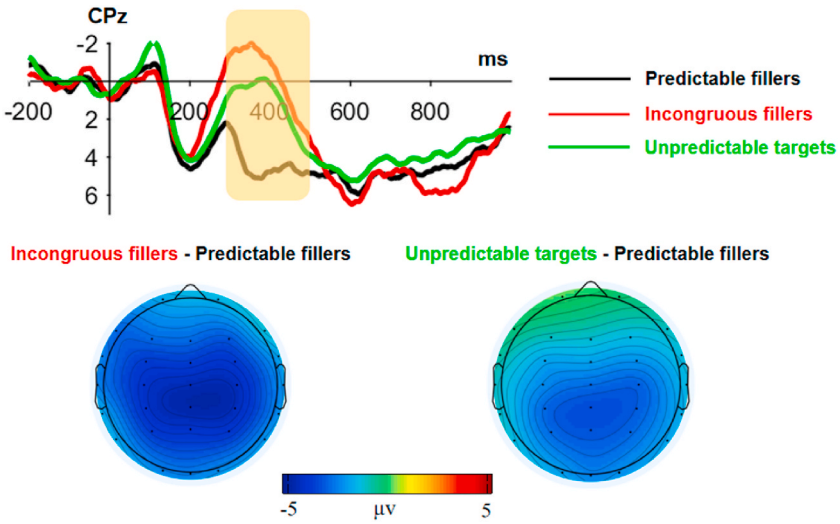


Fig. 3. Grand average ERPs to predictable fillers (black), incongruous fillers (red), and unpredictable targets (green) at a representative central-posterior electrode (CPz). A 20 Hz low pass filter was applied to the waveforms for illustration purposes. The scalp maps show the topographic distribution of N400 effects (300–500 ms) in the contrast to incongruous fillers minus predictable fillers (left) and unpredictable targets minus predictable fillers (right).

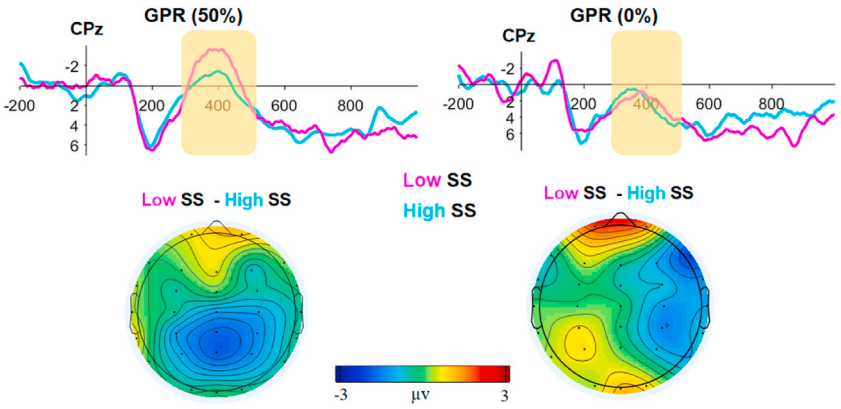


Fig. 4. Grand average ERPs for low SS (magenta) and high SS (cyan) at two levels of GPR. A 20 Hz low pass filter was applied to the waveforms for illustration purposes. The scalp maps show the topographic distribution of the N400 effect (300–500 ms) for low SS minus high SS at 50% GPR level (left) and 0% GPR level (right). GPR = global prediction reliability; SS = semantic similarity.

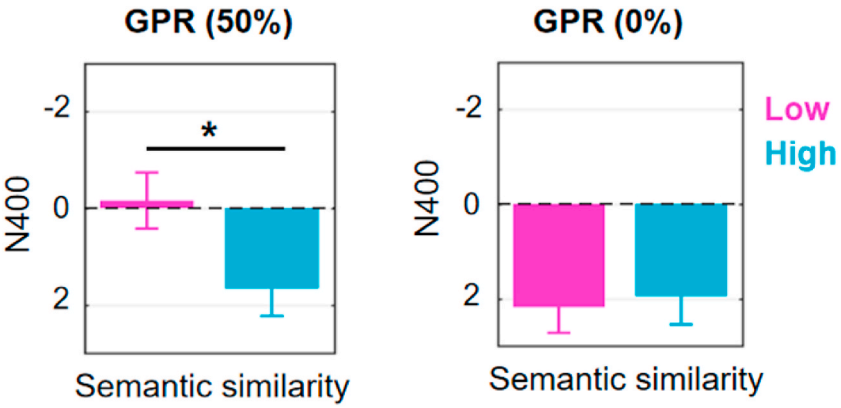


Fig. 5. The bar diagrams show the statistical results of the N400 between high and low semantic similarity groups for two levels of GPR. The left and right panels of the figure show the item-level N400 in predictable (GPR = 50%) and incongruous (GPR = 0%) filler type, respectively. GPR = global prediction reliability. *$p < 0.05$.

(see supplementary materials for details). These results are consistent with a number of previous studies (Kutas & Federmeier, 2010; Zhu et al., 2019) and validate the present EEG data and analysis procedure.

As shown in Fig. 4, we illustrated the waveforms of grand average ERPs for low and high semantic similarity at two levels of GPR (see supplementary materials for waveforms of all scalp electrodes). We also illustrated the scalp maps of the N400 effect (low SS minus high SS) at two levels of GPR. We can see that at the 50% GPR level, low SS showed an increased (more negative) N400 compared with high SS. The N400 effect was mainly distributed in the central-posterior regions and was absent at the 0% GPR level.

The two-by-two ANOVA of the N400 effect within the central-posterior AOI was conducted. The results showed a significant main effect of GPR, $F(1, 39) = 5.75$, $p < 0.05$, $\eta^2 = 0.13$. The N400 amplitude was increased (more negative) at the 50% GPR level than at the 0% GPR level. More importantly, the interaction of GPR by semantic similarity was significant, $F(1, 39) = 4.69$, $p < 0.05$, $\eta^2 = 0.11$. Simple effect analyses showed that the N400 amplitude was reduced (less negative) in the high than low semantic similarity group at 50% GPR level, $t = 2.25$, $p < 0.05$. At the 0% GPR level, the results showed no significant difference between the high and low semantic similarity groups, $t < 1$, $p > 0.7$ (see Fig. 5).

As shown in Fig. 6, we used scatter diagrams to illustrate the correlation between item-level N400 amplitudes and semantic similarity at two levels of GPR while treating semantic similarity as a continuous variable. The results of correlation analyses clearly showed a trend at the 50% GPR level: the N400 amplitude was reduced (less negative) with increasing semantic similarity. This trend was significant, $r = 0.25$, $p < 0.05$. At 0% GPR level, the results showed no significant correlation or clear trend between the N400 and semantic similarity, $r = -0.01$, $p > 0.9$. These results are consistent with previous results reported in Fig. 5.

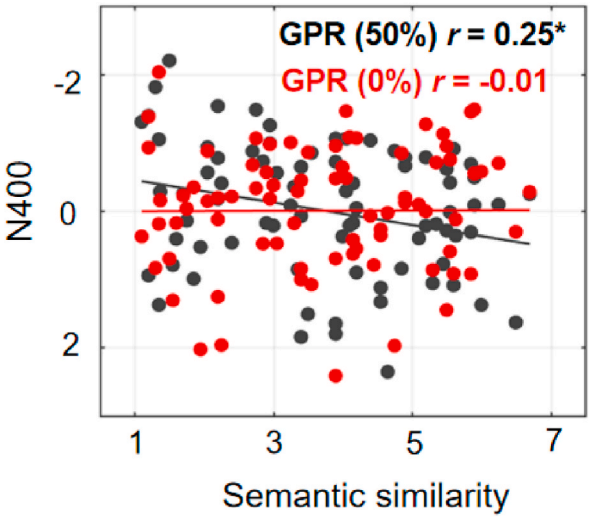


Fig. 6. Correlations of item-level N400 amplitudes against semantic similarity in scatter plots. The black and red circles denote the item-level N400 in the predictable (GPR = 50%) and incongruous (GPR = 0%) filler type, respectively. Best-fitting regression lines are also plotted. GPR = global prediction reliability. *$p < 0.05$.

4. Discussion

The present study examined whether the prediction of semantic features could be adaptively regulated by the global context. In particular, we calculated the semantic similarity between the unpredictable target word and the expected word in experimental sentences. Then, the relationship between the N400 and semantic similarity was examined at each level of GPR. The ERP results showed that the inverse association between N400 and similarity was significant in the predictable filler type (GPR = 50%). However, this relationship was absent in the incongruous filler type (GPR = 0%). Altogether, these results suggested that the activation level of semantic features could be modulated by the global communication context. The implications of these findings are discussed in detail below.

The results showed an interaction between GPR and semantic similarity on the N400 response, with significant effects when the GPR was high. These results extend those of previous studies by suggesting that global context modulates not only lexical predictability, but also the prediction of semantic features of an upcoming word. The predictability of a target word was constructed by manipulating sentential contexts in experimental sentences (see Brothers et al., 2019). For example, the target word (court) was predictable in a high-constraining context (Eric sued the taxi driver and took him to **court** …), but unpredictable in a non-constraining context (Eric picked up his friend and took him to **court** …). Therefore, the N400 effect elicited by experimental sentences was based on whether the prediction of the target word (court) was generated before its occurrence. Unlike in previous studies, all experimental sentences were high constraint in the present study. The reduced item-level N400 in the high semantic similarity group reflected the facilitation of semantic access to the target word because of pre-activated semantic features.

We proposed that the detailed mechanism underlying the N400 interaction between GPR and semantic similarity was as follows. At the 50% level of GPR, the semantic features of expected words (*玫瑰*[*roses*], *星星*[*stars*]) were pre-activated by prediction because participants were encouraged to predict, and all the experimental sentences were high constraint (>0.63). Even when the unpredictable target words (*牡丹*[*peonies*], *钞票*[*cash*]) were seen instead of the expected words, the semantic access of the unpredictable target word was facilitated by pre-activated semantic features. This facilitation was more significant for experimental sentences with high semantic similarity (*牡丹*[*peonies*] - *玫瑰*[*roses*]) compared to low semantic similarity (*钞票*[*cash*] - *星星*[*stars*]). Therefore, the item-level N400 became reduced with increasing semantic similarity. In contrast, at the 0% level of GPR, participants were discouraged from predicting the semantic features of expected words (*玫瑰*[*roses*], *星星*[*stars*]) from the start. Therefore, no significant relationship was shown between the N400 and semantic similarity.

The lack of an inverse relationship between N400 and semantic similarity at the 0% GPR level suggested that participants were discouraged from making predictions during sentence processing because of the global context. Previous studies have shown that words with within-category violations elicited larger N400s in low constraint sentences than in high constraint sentences (Federmeier & Kutas, 1999). This suggests that the N400 is sensitive to semantic feature overlap between the within-category violation word and the expected word, even when the expected word is not presented. These results are consistent with those of the present study. However, our results showed that this potential N400 sensitivity was absent at the 0% GPR level. This suggests that when none of the

expected words was encountered in sentences within a block, participants were more inclined to passively wait for the words to appear instead of making predictions. This argument is consistent with evidence that people may not make predictions in some situations (Huettig & Guerra, 2019; Huettig & Mani, 2016).

One innovation of the present design was that both semantic similarity and levels of prediction strength were manipulated and systematically dissociated. That is, prediction strength was manipulated using the GPR induced by filler sentences. Meanwhile, the semantic similarity was orthogonally manipulated in different experimental sentences. By contrast, most previous paradigms manipulated only semantic similarity or only prediction strength. For example, only semantic similarity was manipulated in a research paradigm that compared between-category violations and within-category violations (Federmeier & Kutas, 1999; Zhu et al., 2009). Consider this sentence: *The gardener really impressed his wife on Valentine's Day. To surprise her, he had secretly grown some* ***roses/tulips/palms***. The semantic similarity of the within-category violation (*tulips*) to the expected word (*roses*) was higher relative to the between-category violation (*palms*), but there was no difference in prediction strength relative to the expected word (*roses*) (Federmeier & Kutas, 1999). In studies using a different paradigm that compared low and high constraint sentences, it was difficult to orthogonally manipulate the semantic similarity (Wlotko & Federmeier, 2012). Even though the prediction strength of expected words was stronger in high than low constraint sentences, it was difficult or controversial to define the expected word in low constraint sentences (Lau, Holcomb, & Kuperberg, 2013; Thornhill & Van Petten, 2012).

In the present study, the inverse relationship between the N400 and semantic similarity was modulated by GPR. However, there was no such relationship in the behavioral results (semantic plausibility rating). The reason for this discrepancy may be that these two measures tap into different processes during sentence comprehension. To be specific, because it benefits from the richer measure of real-time processing, the present N400 effect is time-locked to the target words in experimental sentences. This measure of N400 effect is more sensitive to the semantic facilitation of target words underlying the prediction mechanism, which was the focus of our study. In contrast, the semantic plausibility rating was more related to the processing of the whole sentence. That is, participants combined the meaning of each word into a whole sentence meaning. After reading each sentence, they evaluated the overall semantic plausibility of the sentence against their world knowledge and made responses. Thus, GPR is more likely to modulate the activation of the semantic features of a target word rather than the semantic plausibility of sentences.

It is worth noting that there are two important considerations in the present study. First, it is ideal if the effect of sentence context is counterbalanced, by reversing the semantic relatedness of the target words in the high and low semantic similarity conditions. However, it was not done in the present study because it is difficult (if possible) to construct enough sentences while considering all other requirements. Secondly, the trial number per condition was limited in the present study (Keating & Jegerski, 2015). We would like to validate the present results with more rigorous sentence materials and more trials in future studies.

Taken together, the results of the present study provided direct evidence that the N400 effect is sensitive to the prediction of semantic features and this effect could be modulated by GPR. That is, comprehenders can use information from the global context to regulate the prediction of upcoming content such as semantic features.

Author statement

Wenjia Zhang: Conceptualization, Methodology, Writing-Original Draft. **Jie Dong**: Conceptualization, Methodology, Writing-Original Draft. **Xu Duan**: Formal analysis. **Yi Zhang**: Visualization. **Xuefei Gao**: Writing-Reviewing and Editing. **Anna Zhen**: Writing-Reviewing and Editing. **Jie Zhang**: Methodology, Resources. **Hao Yan**: Writing-Review & Editing, Supervision, Funding acquisition.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Social Science Foundation of China (20BYY097). The authors would like to thank David Holman for helpful comments on earlier versions of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jneuroling.2022.101109.

Appendix A. Experimental sentences of low semantic similarity

Sentence frames	Unpredictable target words	Expected words
晚上他在院子里数	钞票	星星
小王享受乡下清新的	绿茶	空气
集邮爱好者李明买了一套	西装	邮票
书法家的手里握着一支	香烟	毛笔
解放军战士教孩子们叠	罗汉	被子
杀人犯昨天在黑市买了一支	牙膏	手枪
年轻演员登上了梦寐以求的	长城	舞台
小兰站在校园门口派发	糖果	传单
张律师赢了这场	球赛	官司
漂亮的空姐准备登上	舞台	飞机
大伟给女友买了一枚	邮票	戒指
老大娘冒着枪林弹雨给红军送	担架	粮食
烟鬼张伟去商场买了一条	项链	香烟
萍萍拎着水筒去海边拾	垃圾	贝壳
感冒的爸爸正在用纸巾擦	眼镜	鼻涕
冬天城市上空漫天飞舞着	风筝	雪花
爷爷早餐后戴上老花镜看	圣经	报纸
书法家的写字台上放着两瓶	牛奶	墨水
语文老师在课上教小学生查	日历	字典
基督教信徒虔诚地翻开	日历	圣经
小狗专心地咬着	椅子	骨头
婷婷给好朋友送了一束	麦子	鲜花
他在贝壳里发现了一粒	弹球	珍珠
海峡两岸正在进行文化	对抗	交流
专家拆除了恐怖分子安装的	架子	炸弹
王璐脖子上戴着一条	哈达	项链
救助队在草地上搭起了	舞台	帐篷
他的研究工作有了新的	赞助	进展
杜甫是一位伟大的	老人	诗人
外面刮大风小明赶紧关了	电视	窗户
小芸到度假村泡	酒吧	温泉
值日生课前擦好了	地板	黑板
牧羊犬联合起来赶走了	小偷	豺狼
妈妈给远游的儿子织	草鞋	毛衣
每天早上振宇会撕掉昨天的	日记	日历
生物课老师拿着枫叶教大家制作	卡片	标本
总经理决定和那家公司打	交道	官司
他拿出信封并贴上了	标签	邮票
演员丽莎正在后台熟悉	服装	台词
小威坐在电脑前敲打着	桌子	键盘

Appendix B. Experimental sentences of high semantic similarity

Sentence frames	Unpredictable target words	Expected words
当初的幼苗长成了	树林	大树
病人都被安置在	诊室	病房
丰收的果农正在卖	果汁	水果
老年人经常去医院量	体重	血压
踏入职场的黄强买了一双	袜子	皮鞋
晴朗的夜空满是	美景	星星
农民通过喷洒农药杀死	蝗虫	害虫
这支钢笔已经没有了	笔头	墨水
主持人有时会忽悠电视机前的	球迷	观众
暴风雨前村庄上空布满	水汽	乌云
冬天的北京下起了	暴雨	大雪
经理在飞机起飞前赶到了	入口	机场
刚学会写字的小孩手握一支	粉笔	铅笔
小明去理发店修剪	胡子	头发
领导在操场上检阅	警犬	士兵
需要冷藏的食物要放进	厨房	冰箱
电影公司为维护版权利益联合打击	走私	盗版
国家加大力度来搞活	教育	经济
抢匪用身体撞碎了大厅的	镜子	玻璃
小明为不亮的电灯更换	开关	灯泡
热情的观众给予了演员	拥抱	掌声
荒芜的稻田里长了很多	害虫	杂草
小明看电视剧时最讨厌插播	音乐	广告
毕业生在忙着写	小说	论文
她站在桥上欣赏周围美丽的	街道	风景
受凉以后妈妈得了	肺炎	感冒
王慧从手电筒里取出	灯泡	电池
被关在门外的小明忘记带	手机	钥匙
收银员小孟正在认真地点	货物	钞票
李静家的水缸养着	乌龟	金鱼
电脑在高速处理实验的	材料	数据
为促进睡眠小红喝了一杯	糖水	牛奶
看着老照片他流下了	汗水	眼泪
端午节奶奶准备了很多	零食	粽子
导演请人来编写	广告	剧本
学校为学生集体注射	点滴	疫苗
他们在火锅里涮	青菜	羊肉
公司对新上市的产品打	折扣	广告
这场地震使很多孩子成为	残疾	孤儿
刘明为空空的新房添置了一批	书籍	家具

References

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia, 135*, Article 107225.

Clark, A. (2012). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 3*(3), 181–205.

Dave, S., Brothers, T., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2021). Cognitive control mediates age-related changes in flexible anticipatory processing during listening comprehension. *Brain Research*, Article 147573.

Delorme, A., & Makeig, S. (2004). Eeglab: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.

Federmeier, K. D., & 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.

Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39*(2), 133–146.

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.

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1055–1069.

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research, 1706*, 196–208.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language Cognition & Neuroscience, 31*(1), 19–31.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass, 2*(4), 647–670.

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition, 37*(1), 1–32.

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

Kutas, M., & Federmeier, K. D. (2010). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Lau, E., Almeida, D., Hines, P. C., & Poeppel, D. (2009). A lexical basis for N400 context effects: Evidence from MEG. *Brain and Language, 111*(3), 161–172.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25*(3), 484–502.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920–933.

Li, N., Niefind, F., Wang, S., Sommer, W., & Dimigen, O. (2015). Parafoveal processing in reading Chinese sentences: Evidence from event-related brain potentials. *Psychophysiology, 52*(10), 1361–1374.

Lopezcalderon, J., & Luck, S. J. (2014). Erplab: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*(1), 213.

Taylor, W. L. (1953). Cloze procedure": A new tool for measuring readability. *Journalism Quarterly, 30*(4), 415–433.

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.

Wang, L., Wlotko, E., Alexander, E., Schoot, L., Kim, M., Warnke, L., et al. (2020). Neural evidence for the prediction of animacy features during language comprehension: Evidence from MEG and EEG Representational Similarity Analysis. *Journal of Neuroscience, 40*(16), 3278–3291.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage, 62*(1), 356–366.

Zhang, W., Chow, W.-Y., Liang, B., & Wang, S. (2019). Robust effects of predictability across experimental contexts: Evidence from event-related potentials. *Neuropsychologia, 134*, Article 107229.

Zhu, Z., Bastiaansen, M., Hakun, J. G., Petersson, K. M., Wang, S., & Hagoort, P. (2019). Semantic unification modulates N400 and bold signal change in the brain: A simultaneous EEG-fMRI study. *Journal of Neurolinguistics, 52*, Article 100855.

Zhu, Z., Zhang, J. X., Wang, S., Xiao, Z., Huang, J., & Chen, H. C. (2009). Involvement of left inferior frontal gyrus in sentence-level semantic integration. *NeuroImage, 47*(2), 756–763.