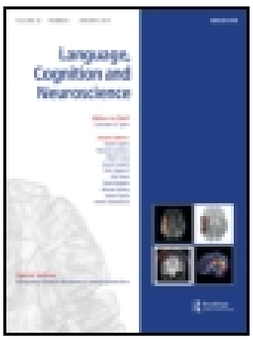

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

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

Wait a second! delayed impact of argument roles on on-line verb prediction

Wing-Yee Chow, Ellen Lau, Suiping Wang & Colin Phillips

To cite this article: Wing-Yee Chow, Ellen Lau, Suiping Wang & Colin Phillips (2018): Wait a second! delayed impact of argument roles on on-line verb prediction, Language, Cognition and Neuroscience

To link to this article: https://doi.org/10.1080/23273798.2018.1427878

View supplementary material

Published online: 24 Jan 2018.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=plcp21

LANGUAGE, COGNITION AND NEUROSCIENCE, 2018
https://doi.org/10.1080/23273798.2018.1427878

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Check for updates

Wait a second! delayed impact of argument roles on on-line verb prediction

Wing-Yee Chow [a], Ellen Lau[b], Suiping Wang[c] and Colin Phillips[b]

[a]Department of Linguistics, University College London, London, UK; [b]Department of Linguistics, University of Maryland College Park, Maryland, USA; [c]Department of Psychology, South China Normal University, Guangzhou, People's Republic of China

ABSTRACT

Comprehenders can use rich contextual information to anticipate upcoming input on the fly, but recent findings suggest that salient information about argument roles may not impact verb prediction. We took advantage of the word order properties of Mandarin Chinese to examine the time course with which argument role information impacts verb prediction. We isolated the contribution of argument role information by manipulating the order of pre-verbal noun phrase arguments while holding lexical information constant, and we examined its effects on accessing the verb in long-term semantic memory by measuring the amplitude of the N400 component. Experiment 1 showed when the verb appeared immediately after its arguments, even strongly constraining argument role information failed to modulate the N400 response to the verb. An N400 effect emerged in Experiment 2 when the verb appeared at a greater delay. Experiment 3 corroborated the contrast between the first two experiments through a within-participants manipulation of the time interval between the arguments and the verb, by varying the position of an adverbial phrase. These results suggest time is a key factor governing how diverse contextual information contributes to predictions. Here argument role information is shown to impact verb prediction, but its effect is not immediate.

ARTICLE HISTORY
Received 14 April 2017
Accepted 5 December 2017

KEYWORDS
Language comprehension; prediction; thematic relations; N400

Introduction

Much recent work suggests that generating predictions about the future is a fundamental principle underlying the brain's operations (Bar, 2011; Hawkins & Blakeslee, 2004; Llinás, 2002). After an explosion of work on this topic across the last decade, predictive processes have now been repeatedly demonstrated across domains such as visual and auditory perception (Bar, 2007; Bendixen, Schroger, & Winkler, 2009), motor planning (Wolpert, 1997), and language comprehension (e.g. Delong, Urbach, & Kutas, 2005; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004). Just as a tennis player's ability to anticipate a flying ball's trajectory is critical to his/her chances of hitting the ball, the ability to anticipate properties of upcoming input ahead of time is likely key to the efficiency with which the brain processes information. Therefore, a better understanding of how predictions are computed is important for understanding the brain's capacity to process information quickly and robustly.

However, recent work in language comprehension has illustrated a notable case in which salient and reliable contextual information appears not to impact predictions. The likelihood of a verb depends not only on the arguments, but also the roles they take. For example, given two arguments "cop" and "thief", the likelihood of the verb "arrest" can differ greatly depending on which argument takes the agent vs. patient role. A verb like "arrest" may be likely if "cop" takes an agent role and "thief" takes a patient role, but it would have been much less likely if the opposite were true. However, some evidence suggests that this information may not be used by humans when they generate predictions online. The goal of the current work is to determine the cause for this puzzling finding, and more broadly, their implications for models of prediction.

Prediction in language comprehension

In the domain of language, previous research has provided much evidence for human comprehenders' ability to generate linguistic predictions on the fly (for a review see Kamide, 2008 ; Kutas, DeLong, & Smith, 2011). For example, listeners are more likely to direct their eye-gaze to a picture of an edible object (e.g. a cake) when they hear the beginning of an utterance like "*The boy will eat … "* compared to a neutral utterance such as "*The boy will move … "* (Altmann & Kamide, 1999). Further, much evidence has suggested that

CONTACT Wing-Yee Chow wingyee.chow@ucl.ac.uk
Supplemental data for this article can be accessed http://dx.doi.org/10.1080/23273798.2018.1427878
© 2018 Informa UK Limited, trading as Taylor & Francis Group

comprehenders compute rich expectations about upcoming inputs at multiple levels of representation (syntactic: Ilkin & Sturt, 2011; Lau, Stroud, Plesch, & Phillips, 2006; Levy, Fedorenko, Breen, & Gibson, 2012; Omaki et al., 2015; Staub & Clifton, 2006; Van Berkum et al., 2005; Wicha et al., 2004; Yoshida, Dickey, & Sturt, 2013; lexico-semantic: Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Otten & Van Berkum, 2008; Szewczyk & Schriefers, 2013; phonological and orthographic: Delong et al., 2005; Dikker, Rabagliati, Farmer, & Pylkkanen, 2010; Dikker, Rabagliati, & Pylkkänen, 2009; Farmer, Yan, Bicknell, & Tanenhaus, 2015; Kim & Lai, 2012; Laszlo & Federmeier, 2009). Here, we operationally define "prediction" as the pre-activation of stored representations before the bottom-up input is encountered, and we will make no a priori assumptions regarding the nature of the mechanisms involved (e.g. whether they are automatic or controlled).

Event-related brain potentials (ERPs) provide a useful tool for investigating the nature of prediction in real time language comprehension. In particular, the amplitude of the N400, a negative-going ERP component that starts at around 250 ms and peaks at around 400 ms post stimulus onset (Kutas & Hillyard, 1980), may be used as a diagnostic of comprehenders' lexical semantic predictions. An N400 response is elicited by any content word (e.g. nouns, verbs, adjectives), whether presented in isolation or in sentences (Kutas & Federmeier, 2011) and it has been proposed to reflect the ease of accessing a word's meaning from long-term memory (Kutas & Federmeier, 2000; Lau, Phillips, & Poeppel, 2008). Crucially, its amplitude is modulated by a word's predictability, which is commonly operationalised as the proportion of trials on which speakers continue the sentence context with that word in an untimed sentence fragment completion task (also known as *cloze probability*; Taylor, 1953). Many have observed an inverse relationship between the amplitude of the N400 response to a word and that word's cloze probability (e.g. Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Gunter, Stowe, & Mulder, 1997; Kutas & Hillyard, 1984; Wlotko & Federmeier, 2012), which has been taken to reflect facilitated lexical semantic access for more predictable words. Under this view, reduction in the N400 response to a given word may be taken to reflect the extent to which that word's meaning has been pre-activated (the "prediction view" of the N400; Federmeier & Kutas, 1999).[1]

N400 amplitude measures have provided some of the primary evidence that comprehenders can rapidly integrate various sources of contextual information to generate predictions. Many factors that are known to affect a word's cloze probability (e.g. negation,[2] sentence structure, event knowledge, world knowledge, message-level representations) have also been shown to modulate N400 amplitude (Bicknell, Elman, Hare, McRae, & Kutas, 2010; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Kos, Vosse, van den Brink, & Hagoort, 2010; Otten, Nieuwland, & Van Berkum, 2007; Paczynski & Kuperberg, 2012; Van Berkum, 2009; Xiang & Kuperberg, 2015). For example, listeners presented with an utterance like "*Every evening I drink some …* " show a smaller N400 response to the word "wine" if the speaker is an adult compared to when it is a child (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008).

It should be noted that the N400 has also been taken to reflect the ease with which a word can be semantically integrated in a given context (the "integration view" of the N400; e.g. Brown & Hagoort, 1993; Van Berkum, Hagoort & Brown, 1999). This view is most commonly associated with the generalisation that semantically or pragmatically incongruous words tend to elicit a larger N400 response than congruous words (e.g. Hagoort et al., 2004). Under this view, the N400's sensitivity to a word's predictability does not arise from pre-activation per se, but merely reflects the fact that words that are more predictable can be semantically integrated more easily. While it is not the primary goal of the present study to resolve this long-standing debate about the functional significance of the N400, we will consider both the prediction and integration views of the N400 and discuss the present findings' potential implications for both accounts.

Prediction in the processing of thematic relations

Previous research using different experimental techniques has demonstrated that comprehenders can use various sources of information to compute thematic relations predictively (Altmann & Kamide, 1999; Aoshima, Phillips, & Weinberg, 2004; Ferretti, McRae, & Hatherell, 2001; Kamide, Altmann, & Haywood, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005; McRae, Hare, Elman, & Ferretti, 2005; Omaki et al., 2015). For example, McRae et al. (2005) reported that participants were faster to name a verb when it was preceded by a noun that was a typical event participant (e.g. agent, patient) and suggested that comprehenders can use nouns to pre-activate the classes of events in which they typically play a role (see also Hare, Jones, Thomson, Kelly, & McRae, 2009). In another study, Kamide et al. (2003) used a visual world eye-tracking paradigm and found that listeners of Japanese, a verb-final language, were more likely to look towards a picture of a hamburger prior to the verb when the first argument (waitress) was marked with nominative case

and the second argument (customer) was dative-marked, compared to when the second argument was accusative-marked. Since nominative, accusative and dative cases tend to mark arguments that serve the agent, patient and goal roles for most verbs in Japanese, these results suggested that listeners can use case markings on pre-verbal arguments to anticipate properties of the upcoming verb (e.g. the number of arguments it takes). Meanwhile, other evidence suggests that comprehenders' predictions about an upcoming verb are also sensitive to information such as the number of preverbal arguments (Brusini, Brun, Brunet, & Christophe, 2015) and the animacy of likely arguments (Altmann, 1999).

However, evidence for the impact of the arguments' structural roles, i.e. which argument is the subject and which is the object, on comprehenders' verb predictions has remained surprisingly scarce. In many languages the arguments' structural roles are unambiguously encoded by word order, e.g. "John likes Mary" vs. "Mary likes John". This information is crucial for interpreting thematic relations in a sentence, and can drastically change the likelihood of an upcoming verb in sentences with a verb-final word order. For example, in a verb-final *ba*-construction sentence in Mandarin Chinese like (1a), the verb "arrest" is much more likely when "cop" is the subject and "thief" is the object than if the same pair of arguments are reversed, as in (1b).

(1) Argument role reversal in a verb-final sentence in Mandarin Chinese:
(a) *jingcha ba xiaotou zhua-le* …
cop BA thief arrest …
"The cop arrested the thief."
(b) *xiaotou ba jingcha zhua-le* …
thief BA cop arrest …
"The thief arrested the cop."

Since the particle *ba* always follows the subject and immediately precedes the direct object, the arguments' identity and their syntactic roles are evident even before the target verb. Nevertheless, there is little evidence to date that comprehenders' expectations about an upcoming verb are sensitive to this information during real-time comprehension. In fact, a number of studies in languages such as Dutch, Mandarin Chinese, Japanese and English have reported that reversing the preverbal arguments did not modulate comprehenders' N400 response at the verb in pairs like (1), despite the large difference in the predictability of the verb (Chow & Phillips, 2013; Chow, Smith, Lau, & Phillips, 2016; Hoeks, Stowe, & Doedens, 2004; Kolk, Chwilla, van Herten, & Oor, 2003; Oishi & Sakamoto, 2010; van Herten, Chwilla, & Kolk, 2006; van Herten, Kolk, & Chwilla, 2005; Ye & Zhou, 2008; cf. Bornkessel-Schlesewsky et al., 2011). These findings are unexpected and constitute a clear exception to the generalisation that the amplitude of the N400 response to a word is inversely related to its offline predictability or cloze probability.

Relatedly, there is an additional body of studies that examine the effects of role reversals in verb-medial sentences (e.g. "The hearty meal was devouring/devoured … "; Kim & Osterhout, 2005; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Stroud & Phillips, 2012). We suspect that there are close parallels in the mechanisms involved, but we focus here on cases in which the arguments' structural roles are known prior to the verb in order to study how argument role information impact verb predictions.

It should be noted, however, that comprehenders do not simply ignore the arguments' structural roles when they interpret these sentences. Many have shown that comprehenders can readily detect the implausibility that results from the reversal of the arguments' roles (e.g. Kolk et al., 2003). Further, ERP studies across different languages have shown that argument role reversals elicit a larger late positivity (a P600 effect; e.g. Kim & Osterhout, 2005; Kolk et al., 2003; Stroud & Phillips, 2012; Ye & Zhou, 2008), which has been associated with error detection and reanalysis (Coulson, King, & Kutas, 1998; Hagoort, Brown, & Groothusen, 1993; Hahne & Friederici, 1999; Kuperberg, 2007; Osterhout & Holcomb, 1992; van de Meerendonk, Kolk, Vissers, & Chwilla, 2010). These results suggest that information about the arguments' structural roles can immediately impact comprehenders' interpretation, but it does not seem to affect the cognitive processes that underlie the N400 response.

Three kinds of accounts have been proposed to explain the N400's surprising insensitivity to argument role reversals. First, some researchers adopted the integration view of the N400 (e.g. Hagoort et al., 2004) and have taken these results to indicate that comprehenders fail to detect implausibility in role-reversed sentences at least initially[3] ("Semantic Illusion" hypothesis, e.g. Bornkessel-Schlesewsky & Schlesewsky, 2008; Hoeks et al., 2004; Kim & Osterhout, 2005; Kolk et al., 2003; Kuperberg, 2007). For example, Kolk et al. (2003) proposed that comprehenders have a bias to compute verb-argument relations based on what is most plausible rather than what the syntax indicates, and thus that upon encountering the verb they compute a plausible interpretation in both the canonical and role-reversed sentences. Under this view, argument role reversals should not modulate the N400 amplitude regardless of the impact on the

verb's predictability, since the verb can always form a plausible thematic relation with the arguments.

On the other hand, some have focused on the N400's sensitivity to lexical semantic association and have attributed the lack of an N400 effect to argument role reversals to the fact that the verb tends to be strongly associated with its arguments in both canonical and role-reversal sentences (Brouwer, Fitz, & Hoeks, 2012). Under this view, argument role reversals should fail to modulate the N400 only when the verb is strongly associated with the arguments, and they should elicit an N400 effect when the verb is only weakly associated with the arguments.

Lastly, with the linking hypothesis that the N400 is modulated by the extent to which a word's semantic representation has been pre-activated (i.e. the prediction view), we have proposed that the N400's insensitivity to argument role reversals may be taken to show that verb predictions are not always sensitive to argument role information, potentially because the impact of argument role information on online verb prediction is not immediate (Chow & Phillips, 2013; Chow, Smith, et al., 2016). We will call this the "Slow Prediction" hypothesis.

In the case of argument role reversals (e.g. in (1)), in order for a verb (e.g. "arrest") to become differentially pre-activated and elicit an N400 effect, comprehenders must not only pre-activate events that likely involve cops and thieves as participants, but those that specifically involve cops as agents and thieves as patients in one case and the opposite in the other. Therefore, even though previous findings suggest that comprehenders can quickly (i) use case markings on pre-verbal arguments to anticipate the properties of the verb (e.g. Kamide et al., 2003) and (ii) use nouns to pre-activate events that typically involve them as participants (e.g. McRae et al., 2005), it is conceivable that comprehenders may require more time to preferentially pre-activate the events that involve those arguments in specific thematic roles.

In a recent paper, we discussed different cognitive mechanisms that may be involved in verb prediction and discussed why argument role information may not have an immediate impact (Chow, Momma, Smith, Lau, & Phillips, 2016). We conceptualised the process of verb prediction as a memory retrieval problem, in which comprehenders use information about the event participants to query their event knowledge in long-term memory in order to predictively identify which events are most likely to be described by the sentence. Further, we proposed that our event knowledge may not be represented in such a way that can be directly queried using linguistically encoded argument role retrieval cues, and thus additional processes may be required for reconfiguring the cues (or the event representations themselves) in order for likely events to be successfully identified and retrieved from memory.

In fact, one common characteristic among many previous studies that looked at argument role reversals is the close proximity between the arguments and the target verb in the experimental materials. For example, the SOA between the second argument and the target verb was only 480 ms in Hoeks et al. (2004), 530 ms in Chow, Smith, et al. (2016), 600 ms in Chow and Phillips (2013), and 645 ms in Kolk et al. (2003) and van Herten et al. (2005, 2006). Therefore, it is possible that comprehenders do use argument role information to predict an upcoming verb, but they simply cannot do so quickly enough when the verb appears immediately following the arguments. This idea is consistent with previous findings that timing manipulations can modulate predictability effects (e.g. Dambacher et al., 2012; Ito, Corley, Pickering, Martin, & Nieuwland, 2016; Kutas, 1993; Wlotko & Federmeier, 2015). Under this view, the N400's sensitivity to argument role reversals should depend on (i) the predictability of the verb and (ii) the amount of time available for comprehenders to incorporate information about the arguments' roles in their verb predictions.

One piece of suggestive evidence for the Slow Prediction hypothesis comes from a visual-world eye-tracking study by Kukona and colleagues, who examined listeners' eye fixations on displays containing verb-related patients (e.g. crook) and agents (e.g. policeman) as they listened to simple English sentences (Kukona, Fang, Aicher, Chen, & Magnuson, 2011). The authors found that listeners were nearly as likely to look at agents as patients when they heard the beginning of active sentences like "Toby arrests the … " (Experiment 1), but they were more likely to look at agents when they heard the beginning of passive sentences such as "Toby was arrested by the … " (Experiment 2). With the linking hypothesis that listeners' anticipatory eye-movements reflect their online predictions, these results may be taken to show that listeners' predictions are sensitive to filled vs. unfilled thematic roles in passive sentences but not in active sentences, potentially because they had slightly more time to predict the second argument in passive sentences than in active sentences (due to the presence of additional words "was" and "by" in passive sentences, although with spoken stimuli the duration difference may be small). However, as the authors acknowledged, passive and active sentences differed in other ways (e.g. the nature and number of the syntactic cues available), and anticipatory fixations may not be fully reflective of comprehenders' predictions as they can also be driven by priming (e.g. Huettig & Altmann,

2005; Yee & Sedivy, 2006). Further, the fact that the target was always presented with a small set of distractors in the visual world paradigm may also artificially restrict (or even interfere with) comprehenders' predictions. Therefore, even though these results are suggestive they should not be taken as evidence for the hypothesis that argument role information has a delayed impact on predictions.

In the current study we tested competing accounts for the N400's insensitivity to argument role reversals by examining the effects of argument role reversals under different circumstances in three ERP experiments. We used the highly frequent *ba*-construction in Mandarin Chinese to create canonical and role-reversed sentences (as illustrated in 1a-b above). In Experiments 1 and 2, argument role information was manipulated by reversing the order of the preverbal arguments in sentences in which the verb had high vs. low predictability in the canonical condition, and we controlled the timing of the verb relative to the arguments by manipulating between experiments the linear distance between the arguments and the verb. The verb immediately followed the arguments in Experiment 1 (e.g. "cop *ba* thief arrest"); a temporal phrase was placed between the arguments and the verb to increase the linear distance between them in Experiment 2 (e.g. "cop *ba* thief (yesterday afternoon) arrest"). In Experiment 3 we manipulated the linear distance between the arguments and the verb within the same experiment.

Experiment 1

In this experiment we examined whether argument role reversals impact the N400 response to the verb when they have a large and demonstrable impact on the verb's cloze probability. Previous studies that examined the effects of argument role reversals did not measure cloze probability, because they typically focused on how reversals create implausible sentences, rather than on how they change cloze probabilities. Therefore, it is unclear to what extent the target verb's cloze probability was modulated by argument role reversals in those studies. Here we explicitly measured and manipulated the impact of argument role reversals on the target verb's cloze probability. In *high-predictability sentences* the target verb had high cloze probability in the canonical control condition; in *low-predictability sentences* the verb had low cloze probability in the canonical condition. In both cases the verbs always had zero cloze probability when the arguments were reversed. Therefore, argument role reversals resulted in a much larger cloze probability difference between the canonical and role-reversed conditions in the high-predictability sentences (64% vs. 0%) than in the low-predictability sentences (7% vs. 0%). Sentences were plausible in the canonical condition and implausible in the role-reversed condition. Since each pair of canonical and role-reversed sentences contained identical sets of words, argument role reversals effectively isolated the contribution of the arguments' structural roles while holding other factors such as lexical semantic association constant.

Therefore, if the N400 is sensitive to the verb's cloze probability and its insensitivity to argument role reversals is attributable to a lack of cloze probability differences in previous studies, then argument role reversals should elicit an N400 effect when there is a clear cloze probability difference between the canonical and role-reversed conditions (i.e. in the high-predictability sentences). Alternatively, if argument role reversals fail to elicit an N400 effect because (i) argument role information cannot impact comprehenders' verb predictions quickly enough (Slow Prediction hypothesis), or (ii) comprehenders temporarily fail to detect the implausibility in role-reversed sentences (Semantic Illusion hypothesis), then no N400 effects should be observed in high- and low-predictability sentences alike.

Methods

Participants

Twenty-four students (19 female, mean age = 21 years, range 18–24 years) from South China Normal University participated in the current study. All participants were native speakers of Mandarin Chinese, were strongly right-handed based on the Edinburgh Handedness Inventory (scoring 90% or above; Oldfield, 1971), and had normal or corrected-to-normal vision and no history of neurological disorder. None of the participants in the present study took part in more than one of the experiments. All participants gave informed consent and were paid 20 RMB/hour for their participation.

Cloze probability norming

Cloze probability estimates were gathered from 60 student volunteers at South China Normal University on a total of 190 pairs of subject-*ba*-object sentence frames (e.g. "*cop ba* thief … " and its role-reversed counterpart "thief *ba* cop … "). None of these participants took part in any of the ERP experiments. The sentence frames were divided into two lists, and each list was completed by 30 participants. Participants were asked to provide the best continuations for the sentence frames with one or more words. Overwhelmingly participants responded with a verb which completed the sentence. A few exceptions were found when participants extended the second NP before supplying a

verb (e.g. assistant *ba* manager's *speech* recorded) or preceded the verb with a preposition phrase (e.g. magician *ba* pigeon *from a hat* pulled out); we always used the first word following the sentence fragment to compute cloze counts.

Materials

Across Experiments 1–3, all experimental sentences used the highly frequent SOV *ba*-construction in Mandarin Chinese. Role-reversed sentences were created by reversing the order of the pre-verbal arguments in the canonical sentences. Within each item set the canonical and role-reversed sentences had an identical verb-argument triplet and differed only in the order of the arguments. Both of the arguments were animate and the order of the arguments alone did not induce any anomaly prior to the target verb. Further, in order to avoid sentence-final wrap-up effects the sentences were extended beyond the target verb with words that were held constant across conditions within each item set.

The experimental stimuli consisted of 120 pairs of sentences, each with a canonical and reversed argument order. Argument order was manipulated in two groups of sentences (Table 1). In the *high-predictability sentences* the target verb had an average cloze probability of 64% (range 41%–97%) in the canonical condition; in the *low-predictability sentences* the verb had an average cloze probability of 7% (range 3%–21%) in the canonical condition. In both cases the verbs always had zero cloze probability when the arguments were reversed.[4] High- and low-predictability sentences shared the same SOV *ba*-construction but contained different arguments and verbs. All sentences were plausible in the canonical conditions and implausible in the role-reversed conditions. A complete set of the experimental materials from the present study (Experiments 1–3) is available in the Supplementary Materials.

Table 1. Experimental conditions and sample materials in Experiment 1 (top) and sample materials for the control comparison in Experiments 1–3 (bottom).

Experimental comparison

Low-predictability

Canonical: 老刘 把 鹦鹉 / Mr. Liu BA parrot — 训练了 好一段时间 。 **train** quite some time.

Role-reversed: 鹦鹉 把 老刘 / Parrot BA Mr. Liu

"Mr. Liu trained the parrot for quite some time."
vs. "The parrot trained Mr. Liu for quite some time."

High-predictability

Canonical: 警察 把 小偷 / cop BA thief — 抓了 回警局。 **arrest** (and bring back) to police station.

Role-reversed: 小偷 把 警察 / thief BA cop

"The cop arrested the thief (and brought him back) to the station."
vs. "The thief arrested the cop (and brought him back) to the station."

Control comparison

(Identical across Experiments 1–3)

Expected: 牧民 从 羊身上 剪下 很多 毛 …
Herdsman from sheep's body cut-down a lot of **wool** …
"From the sheep the herdsman collected a lot of wool … "

Unexpected: 老板 在 去年 收购 很多 毛 …
Boss ZAI last year buy a lot of **wool** …
"Last year the boss bought a lot of wool … "

In addition, across Experiments 1–3, 60 plausible filler sentences from a previous study were included to examine the effects of a standard cloze probability manipulation in the populations studied here (a control comparison; bottom of Table 1). In each item, the same target word was paired with two different sentence contexts, such that the target word had a high cloze probability (M = 88%, SD = 11%) in one and a low cloze probability (M = 27%, SD = 13%) in the other. A group of 48 participants were asked to rate the plausibility of these sentences on a 5-point scale. Both high and low cloze sentences were rated as highly plausible (high cloze: 4.3; low cloze: 4.1). The position of the target word was matched across conditions in each set. As was the case in many previous studies, the target word was preceded by completely different contexts across conditions. Therefore, all of the words in the sentence context could have contributed to the differential predictability of the target word in these sentences. This critically differed from the experimental sentences, in which the target verb was preceded by sentence contexts that only differed by the order of the arguments across conditions.

Experimental sentences were distributed across two presentation lists, such that only one version of each item appeared in each list. Each list contained 120 experimental sentences (30 per condition), 60 plausible filler sentences for the control comparison and 60 unrelated implausible filler sentences of similar length and structural complexity, so that the overall plausible-to-implausible ratio in each presentation list was 1:1, in line with many previous ERP studies (e.g. Hoeks et al., 2004; Kim & Osterhout, 2005).

Procedure

Participants were comfortably seated about 100 cm in front of a computer screen in a testing room. Sentences were segmented into words (see example in (2)), which were presented one at a time in a white font (30 pt simplified Chinese characters) on a black background at the centre of the screen. Each sentence was preceded by a fixation cross that appeared for 500 ms. Each word appeared on the screen for 400 ms, followed by 200 ms of blank screen. Each word consisted of 1–4 characters. The last word of each sentence was marked with a

period, followed 1000 ms later by a response cue "?". Participants were instructed to avoid eye blinks and movements during the presentation of the sentences, and they were asked to read each sentence attentively and to indicate whether the sentence meaning was plausible by pressing one of two buttons. Prior to the experimental session, participants were presented with 6 practice trials with feedback to familiarise themselves with the task. The experimental session was divided into four blocks of 60 sentences each, with short pauses in between. Including set-up time, an experimental session lasted around 1.5 h on average.

(1) Sentence segmentation for stimulus presentation:
/去年初，/老刘/把/鹦鹉/训练了/好一段/时间 。/
/Beginning of last year,/Mr. Liu/BA/parrot/train/(for) quite some/time./

EEG recording

EEG was recorded continuously from 30 AgCl electrodes mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1/2, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P3/4, P7/8, and O1/2. The electro-oculogram (EOG) was recorded at four electrode sites; vertical EOG was recorded from electrodes placed above and below the left eye and the horizontal EOG was recorded from electrodes situated at the outer canthus of each eye. Electrode impedances were kept below 5 kΩ. The EEG and EOG recordings were amplified (band-pass filtered at DC - 200 Hz) and digitised online at 1 kHz. Online recordings were referenced to the left mastoid. They were re-referenced to the average of the left and right mastoids and filtered using a 0.1–40 Hz band-pass filter offline. This high-pass filter value was tested in a recent study and found to not introduce artefactual effects (Tanner, Morgan-Short, & Luck, 2015). A 15 Hz low-pass filter was applied for the presentation of grand averages only, but was not used in the analyses.

ERP data analysis

Event-related potentials were computed separately for each participant and each condition for the 1000 ms after the onset of the target verb relative to a 100 ms baseline preceding the target verb. ERP data in all trials (regardless of end-of-trial behavioural response) were evaluated for EOG and other artefacts. Trials contaminated by artefacts were excluded from the averaging procedure. This affected 7.8% of experimental trials, roughly equally distributed across conditions (ranging between 6.3 and 9.6% across conditions).

Statistical analyses on average voltage amplitudes were conducted separately for two time windows chosen based on previous literature and on visual inspection of the data: 350–450 ms for the N400, and 600–800 ms for the P600.

Data from the experimental sentences were analysed using two parallel approaches. As our hypotheses centred on the effect of role-reversal on the N400 and P600, our primary analysis examined the effect of role reversal on these ERP components within each level of predictability. Since both N400 and P600 effects tend to be broadly distributed with a (central-)posterior maximum, we chose a cluster of posterior electrodes for this analysis so as to use the same electrodes to capture both N400 and P600 effects (see also Ito, Corley, et al., 2016). Further, in order to allow a more comprehensive evaluation of effects across the scalp, we also conducted omnibus ANOVAs at each time window to test for main effects and interactions with distributional factors.

For the posterior electrode cluster analyses, we averaged participants' ERPs across 6 posterior electrodes (P3, PZ, P4, O1, OZ, and O2) and ran paired-sample t-tests to examine the effect of role reversal within each level of predictability in each time window. Further, we used the *bootES* package (Gerlanc & Kirby, 2013) to compute standardised effect sizes of these contrasts (Pearson's *r*) and their bootstrap confidence intervals (95% CI). An advantage of using Pearson's *r* to express the effect size of a contrast is that the effect is scaled to have an absolute magnitude between 0 and 1.

For the omnibus repeated measures ANOVAs, we examined time-window average ERPs at 18 electrodes, fully crossing predictability (high vs. low) and reversal (congruous vs. role-reversed) with anteriority (anterior vs. central vs. posterior) and laterality (Left vs. Midline vs. Right). We assigned topographical grouping factors to electrodes as follows: left-anterior: F3, FC3; midline-anterior: FZ, FCZ; right-anterior: F4, FC4; left-central: C3, CP3; midline-central: CZ, CPZ; right-central: C4, CP4; left-posterior: P3, O1; midline-posterior: PZ, OZ; right-posterior: P4, O2. Univariate F-tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959). Since predictability was manipulated between different item sets (i.e. high and low-predictability sentences were completely different), main effects of predictability are not meaningful and we discuss effects of predictability only when they interact with the effects of reversal. We also conducted a predictability × anteriority × laterality repeated measures ANOVA in each time window for the filler sentences to examine the effect of cloze probability.

Results

Plausibility judgments

Across both high- and low-predictability sentences, participants judged canonical sentences to be plausible and role-reversed sentences to be implausible with an overall accuracy of 90.6% (Low predictability, canonical = 88.3%; Low predictability, role-reversed = 90.4%; High predictability, canonical = 90.3%; High predictability, role-reversed = 93.3%).

Event-related potentials (Experimental comparisons)

The top half of Figure 1 shows the grand average ERPs at centro-posterior electrode CPZ and the topographic distribution of the effects of argument role reversals at the target word in the 350–450 ms and 600–800 ms time intervals in the low-predictability and high-predictability conditions in Experiment 1. Results of the posterior electrode cluster analyses and the omnibus ANOVAs are presented in Table 2 and the leftmost column of Table 3 respectively. Grand average ERPs across all 30 electrode sites are presented in Supplementary Figures 1 and 2.

As shown in Figure 1, argument role reversals elicited a clear P600 effect with the same topographic distribution across both high- and low-predictability sentences, indicating that participants quickly and reliably detected that the verb was anomalous in the role-reversed conditions. However, there was no clear effect

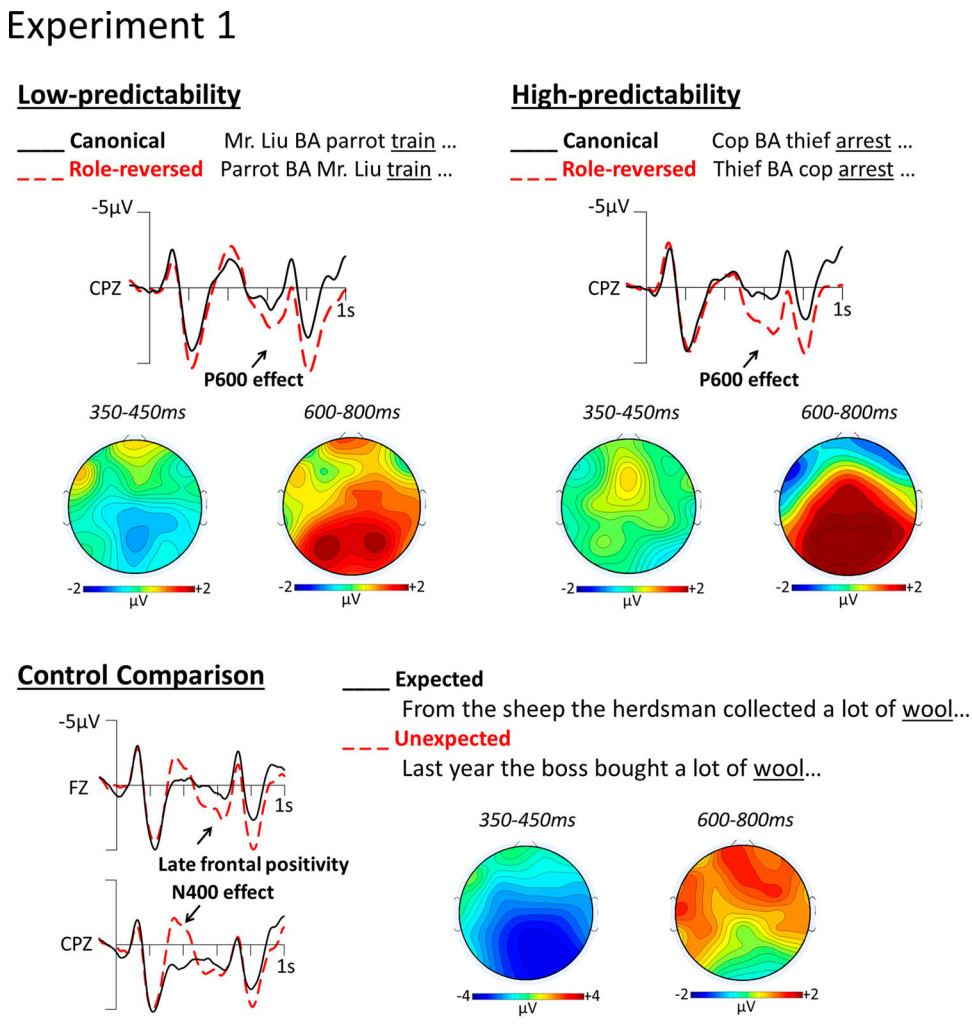

Figure 1. *Top*: Grand average ERPs at centro-parietal electrode CPZ and the topographic distribution of ERP effects in the 350–450 ms and 600–800 ms intervals (role-reversed minus canonical) in the low-predictability (left) and high-predictability (right) conditions in Experiment 1, in which the target verb appeared immediately following the second argument (600 ms SOA). Argument role reversals elicited a P600 effect in both low-predictability and high-predictability sentences, but the N400 was not modulated by the predictability of the target verb in either condition. *Bottom*: Grand average ERPs at frontal electrode FZ and at centro-posterior electrode CPZ and topographic distribution of ERP effects in the 350–450 ms and 600–800 ms intervals (expected minus unexpected) in the control comparison in Experiment 1. A cloze probability manipulation in the filler sentences elicited a clear N400 effect followed by a frontally distributed late positivity.

Table 2. Results of the posterior electrode cluster analyses in Experiment 1: *t* value, standardised effect size (Pearson's *r*) and 95% confidence interval for the contrast between the canonical and role-reversed conditions in each time interval.

Experiment 1	*df*	*T*	Pearson's *r*	95% CI
Low predictability				
350–450 ms	23	1.12	−0.23	[−0.55, 0.18]
600–800 ms	23	3.43**	0.58	[0.24, 0.76]
High predictability				
350–450 ms	23	<1	−0.07	[−0.42, 0.34]
600–800 ms	23	6.4**	0.80	[0.66, 0.87]

**$p < .01$; *$p < .05$.

of reversal on the N400 in either pair of conditions, particularly in the high-predictability conditions where the cloze probability difference was largest.

In the 350–450 ms interval, the posterior electrode cluster analyses revealed no significant effects of reversal in either high- or low-predictability conditions. The omnibus ANOVA revealed a three-way reversal × anteriority × laterality interaction, but further analyses by anteriority and laterality did not reveal any significant effect.

In the 600–800 ms interval, the posterior electrode cluster analyses showed a larger P600 in the role-reversed condition than in the canonical condition in high- and low-predictability sentences alike. The omnibus ANOVA similarly revealed a main effect of reversal, along with reversal × anteriority, reversal × laterality, and reversal × anteriority × laterality interactions, reflecting a larger central-posterior positivity (P600) for the role-reversed condition in high- and low-predictability sentences alike.

Event-related potentials (Control comparison)

The bottom half of Figure 1 shows the grand average ERPs at frontal electrode FZ and at centro-posterior electrode CPZ and the topographic distribution of ERP effects in the control comparison. Results of the statistical analyses are presented in the leftmost column of Table 4.

The standard cloze probability manipulation in the filler sentences elicited a clear N400 effect and a frontally distributed late positivity. Statistical analyses in the 350–450 ms time interval revealed a significant main effect of predictability, showing that expected words elicit a smaller N400 response than unexpected words. Interactions between predictability and the topographic factors indicated that the effect was largest at midline central and posterior sites. Analyses in the 600–800 ms time interval revealed a significant predictability × anteriority interaction, due to the fact that unexpected words elicited a larger frontal positivity than expected words.

As shown in Figure 1, the topographic distribution of this late positivity was distinct from the posteriorly distributed P600 effect in the experimental sentences. This frontally distributed late positivity has been observed in previous studies and has been argued to reflect processes that are triggered when strong predictions are disconfirmed (DeLong, Quante, & Kutas, 2014; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Thornhill & van Petten, 2012; van Petten & Luka, 2012). We will return to discuss the relationship between this late frontal positivity and the P600 in the General Discussion.

Table 3. Omnibus ANOVA *F*-values at the target verb in Experiments 1–3.

	Experiment 1			Experiment 2			Experiment 3		
Omnibus ANOVA	df	350–450 ms	600–800 ms	df	350–450 ms	600–800 ms	df	350–450 ms	600–800 ms
rev	1,23	<1	18.78**	1,23	1.37	19.01**	1,23	<1	17.95**
*rev * type*	1,23	<1	<1	1,23	1.01	<1	1,23	<1	<1
*rev * ant*	2,46	1.24	5.97*	2,46	<1	29.63**	2,46	<1	13.9**
*rev * lat*	2,46	<1	4.19*	2,46	<1	<1	2,46	<1	2.36
*rev * type * ant*	2,46	<1	1.03	2,46	<1	<1	2,46	<1	< 1
*rev * type * lat*	2,46	<1	1.94	2,46	<1	<1	2,46	5.68**	1.25
*rev * ant * lat*	4,92	3.66*	4.89**	4,92	1.74	1.47	4,92	<1	1.25
*rev * type * ant * lat*	4,92	<1	1.2	4,92	1.35	<1	4,92	1.08	<1

Factors: *type* = predictability (Experiments 1 & 2) or distance (Experiment 3); *ant* = anteriority; *lat* = laterality.
**$p < .01$; *$p < .05$.

Table 4. Omnibus ANOVA F-values at the target word for the control comparisons in Experiments 1–3.

	Experiment 1			Experiment 2			Experiment 3		
	df	350–450 ms	600–800 ms	df	350–450 ms	600–800 ms	df	350–450 ms	600–800 ms
Omnibus ANOVA									
pred	1,23	20.3**	2.25	1,23	45.8**	2.09	1,23	6.93*	3.11^
pred * ant	2,46	7.21**	5.18*	2,46	9.02**	4.26*	2,46	10.44**	8.69**
pred * lat	2,46	3.56*	<1	2,46	<1	1.68	2,46	<1	<1
pred * ant * lat	4,92	2.91*	1.67	4,92	3.51*	4.13**	4,92	3.07*	5**

Factors: *pred* = predictability; *ant* = anteriority; *lat* = laterality.
**$p < .01$; *$p < .05$; ^ $.05 < p < .1$.

Discussion

This experiment showed that, when a verb appeared immediately following its arguments, argument role reversals did not modulate comprehenders' N400 response to the verb even when they greatly impacted its offline predictability, as quantified by cloze probability. The N400's insensitivity to the cloze probability difference resulting from argument role reversals is striking, especially in the context of the P600's sensitivity to argument role reversals as well as the N400's sensitivity to cloze probability in the control comparison. This finding is in line with the absence of N400 effects in previous studies that examined the effect of pre-verbal argument role reversals across different languages (e.g. Chow & Phillips, 2013; Chow, Smith, et al., 2016; Hoeks et al., 2004; Kolk et al., 2003; Oishi & Sakamoto, 2010; van Herten et al., 2006, 2005; Ye & Zhou, 2008), although the effect of argument role reversals on the verb's offline predictability was quantified in only one of these studies (Chow, Smith, et al., 2016). Importantly, the observation in the current study that role reversals do not impact the N400, even when they have a large impact on cloze probability, suggests that the N400's insensitivity is not attributable to a weak experimental manipulation. This is a clear exception to the well-established generalisation that the N400 is modulated by a word's cloze probability.

Meanwhile, as in previous studies, participants were highly accurate in their plausibility judgments and they showed a clear P600 effect to argument role reversals. We extended previous findings by showing that the effect of argument role reversal on the P600 is not modulated by the (un)predictability of the verb. This robust sensitivity suggests that comprehenders computed an accurate interpretation of the sentences and readily detected the implausibility resulting from argument role reversals.

Different aspects of the current results allow us to evaluate competing interpretations of this finding. First, the participants in the current study were unlikely to be processing less deeply or predictively than previous studies showing N400 cloze probability effects because they displayed standard N400 sensitivity to a word's cloze probability in the filler sentences.[5] Further, since the average N400 at the targets in the high-predictability sentences was clearly larger than that elicited by the expected words in the filler sentences, we consider it unlikely that argument role reversals failed to modulate the N400 simply because the lexical association between the verb and the arguments was so strong as to yield a floor effect in the N400 amplitude (although we cannot conclusively rule this out).

However, the present results are still compatible with two competing hypotheses. On the one hand, these results are compatible with the Slow Prediction hypothesis as the N400's insensitivity to role reversals may show that argument role information cannot impact comprehenders' verb predictions quickly enough. Meanwhile, they also cannot rule out the Semantic Illusion hypothesis, according to which the N400's insensitivity reflects a brief initial stage when comprehenders failed to detect implausibility (e.g. Bornkessel-Schlesewsky & Schlesewsky, 2008; Hoeks et al., 2004; Kim & Osterhout, 2005; Kolk et al., 2003; Kuperberg, 2007). We address this possibility more directly in Experiments 2 and 3.

Experiment 2

In Experiment 2 we contrasted the Slow Prediction hypothesis with the Semantic Illusion hypothesis. To this end, we employed the same experimental design as Experiment 1 and increased the distance between the arguments and the verb. This was achieved by placing an adverbial time expression (e.g. "*shangxingqi*", last week) along with the preposition "*zai*" between the second argument and the verb. As sentence fragments were presented at a fixed rate (600 ms SOA), this effectively delayed the presentation of the verb relative to its arguments by 1200 ms. As in Experiment 1, each pair of canonical and role-reversed sentences contained identical sets of words and differed only in the order of the pre-verbal arguments.

Under the Slow Prediction hypothesis, if information about the arguments' structural roles can impact comprehenders' verb prediction within the extended time interval, then argument role reversals should elicit an N400 effect in the high-predictability conditions in this experiment. Even though argument role reversals also make the sentence implausible in the low-predictability conditions, we do not expect to observe an N400 effect there because role reversals does not give rise to a clear difference in the verb's predictability low-predictability sentences. Alternatively, if comprehenders simply do not incorporate information about the arguments' structural roles into their verb prediction (or if they fail to do so within the extended time interval), then the N400 should remain insensitive to argument role reversals. Lastly, since both the high- and low-predictability conditions compared plausible and implausible sentences, the Semantic Illusion hypothesis would predict the same pattern of results in the high- and low-predictability conditions in this experiment. Either both should show no N400 effects (if the heuristics is in play) or both should show an N400 effect (if the more careful algorithm has had time to compute).

Methods

Participants

Twenty-four students (19 female, mean age = 19.9 years, range 18–24 years) from South China Normal University participated in the current study. Informed consent was obtained from all participants. All were right-handed, native Mandarin Chinese speakers with normal or corrected-to-normal vision and were paid for their participation.

Materials

The current experiment shared the same predictability × reversal design as Experiment 1. The experimental materials were adapted directly from those in Experiment 1 to increase the linear distance between the target verb and its arguments. A temporal adverbial phrase (e.g. "zai shangxingqi", last week) always appeared between the second argument and the verb (e.g. cop BA thief ZAI last week arrest …), thereby increasing the SOA between them from 600 ms to 1800 ms. We favoured this over an overall increase in the SOA because a small increase in the overall SOA (e.g. to 700 ms) might not provide sufficient time for argument role information to impact comprehenders' verb predictions before the verb appears, whereas a larger SOA increase (e.g. to 1000 ms) would force participants to read unnaturally slowly and could place excessive demands on their working memory. By contrast, introducing a filler phrase between the second argument and the verb can easily add hundreds of milliseconds to the SOA between them without causing participants to read any more slowly.

The choice of filler phrases was based on two considerations. On the one hand, the naturalness of this word order is sensitive to the "heaviness" (or length) of the adverbial phrase, such that sentences like "Jingcha ba xiaotou zai zuotian-xiawu zhua-le … " (cop BA thief ZAI yesterday afternoon arrest …) would sound more natural than "Jingcha ba xiaotou zai zuotian zhua-le" (cop BA thief ZAI yesterday arrest …). On the other hand, introducing a long and complex adverbial might prevent comprehenders from further processing the main clause during the added time interval. Therefore, we chose to use simple time adverbials that are 4–5 characters long here. All sentences were checked by two native Mandarin-speaking informants to ensure their naturalness prior to the study, and plausibility rating data from 12 speakers who did not participate in any of the present experiments showed that canonical sentences were rated high in plausibility in both short-distance (6.4 out of 7) and long-distance (5.7 out of 7) conditions. Although temporal phrases more commonly appear at the beginning of a sentence or following the subject in Mandarin Chinese, it is important to note that comparisons were always made between canonical and role-reversed sentences with identical temporal phrase placement, and so any differences between conditions are unlikely to be due to constraints on adverb placement.

In order to verify that the adverbial phrases did not make the target verbs more predictable, we obtained cloze probability estimates for the target verbs in these adapted sentences from 60 additional students from South China Normal University who did not participate in any of the ERP experiments. The cloze probabilities for the Experiment 2 items closely resembled the cloze probabilities for Experiment 1 items. In Experiment 2 the target verb had an average cloze probability of 57.1% (range = 20%–100%) and 8.7% (range = 0%–40%)[6] in the canonical conditions in the *high- and low-predictability sentences* respectively. In both cases the verbs' cloze probability was close to zero when the arguments were reversed (*high-predictability sentences*: 1.9%; *low predictability sentences*: 0.4%). This confirms that the temporal adverbials did not themselves affect the cloze probability of the target verbs.[7]

As in Experiment 1, the experimental sentences were distributed in two presentation lists of 240 items. Each list contained 120 experimental sentences (30 per condition) and the same set of 120 filler sentences used in Experiment 1.

Procedure, EEG recording and ERP data analysis

The procedures for the experiment, EEG recording and data analysis were identical to Experiment 1. A total of 11.5% of experimental trials, roughly equally distributed across conditions (9%–13.6%), were excluded from the averaging procedure due to artefacts.

Results

Plausibility judgments

Participants reliably judged canonical sentences to be plausible and the role-reversed sentences to be implausible with an overall accuracy of 90.9% (Low predictability, canonical = 88.0%; Low predictability, role-reversed = 89.5%; High predictability, canonical = 90.7%; High predictability, role-reversed = 95.1%).

Event-related potentials (Experimental comparisons)

The top half of Figure 2 shows the grand average ERPs at centro-parietal electrode CPZ and the topographic distribution of the ERP effects in the 350–450 ms and 600–800 ms time intervals in the low-predictability and

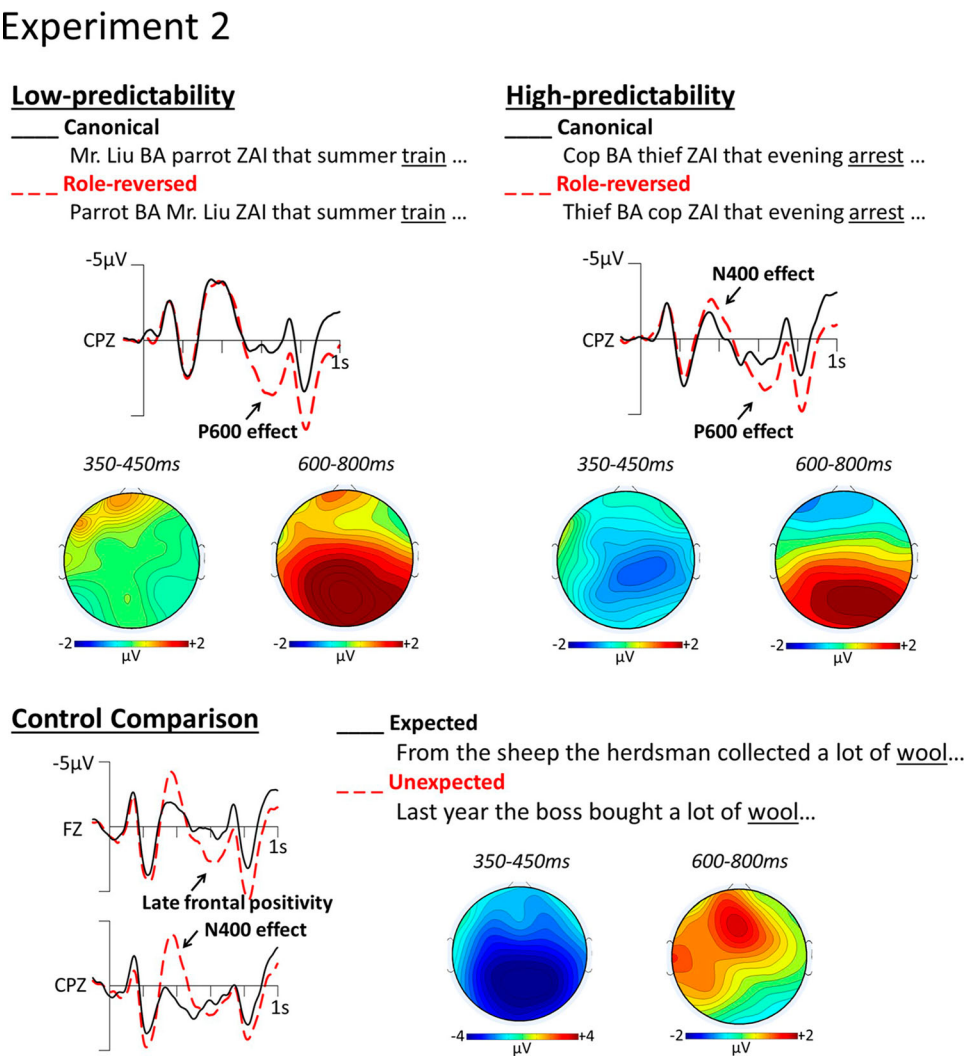

Figure 2. *Top*: Grand average ERPs at centro-posterior electrode CPZ and topographic distribution of ERP effects in the 350–450 ms and 600–800 ms intervals (role-reversed minus canonical) in the low-predictability (left) and high-predictability (right) conditions in Experiment 2, in which the target verb appeared 1800 ms after the onset of the second argument. Argument role reversals elicited a P600 effect in the low-predictability conditions, and an N400 effect followed by a P600 effect in the high-predictability conditions. *Bottom*: Grand average ERPs at frontal electrode FZ and at centro-posterior electrode CPZ and topographic distribution of ERP effects in the 350–450 ms and 600–800 ms intervals (expected minus unexpected) in the control comparison in Experiment 2. A cloze probability manipulation in the filler sentences elicited a clear N400 effect followed by a frontally distributed late positivity.

high-predictability conditions in Experiment 2. Results of the posterior cluster analyses and the omnibus ANOVAs are presented in Table 5 and the middle column of Table 3 respectively. Grand average ERPs across all 30 electrode sites are presented in Supplementary Figures 3 and 4.

Table 5. Results of the posterior electrode cluster analyses in Experiment 2: t value, standardised effect size (Pearson's r) and 95% confidence interval for the contrast between the canonical and role-reversed conditions in each time interval.

Experiment 2	*df*	*t*	Pearson's *r*	95% CI
Low predictability				
350–450 ms	23	< 1	0.01	[-0.42, 0.38]
600–800 ms	23	4.53**	0.69	[0.48, 0.81]
High predictability				
350–450 ms	23	2.24*	-0.42	[-0.65, 0.06]
600–800 ms	23	4.5**	0.68	[0.44, 0.82]

**$p < .01$; *$p < .05$.

As in Experiment 1, a clear P600 effect to argument role reversals was observed in both high- and low-predictability conditions, indicating that participants detected the anomaly in the role-reversed sentences. However, in contrast to Experiment 1, argument role reversals in the high-predictability conditions appeared to also elicit an N400 effect.

In the 350–450 ms interval, the posterior electrode cluster analyses showed that argument role reversals elicited a significant negativity (an N400 effect) in the high-predictability conditions but not in the low-predictability

conditions. The main effect of reversal and the interaction between reversal and predictability failed to reach statistical significance in the omnibus ANOVA, likely because the size of the effect was not very large even in the high-predictability conditions (~0.7 μV; $r = 0.42$).

In the 600–800 ms interval, the posterior electrode cluster analyses revealed a larger P600 in the role-reversed condition than in the canonical condition in high- and low-predictability sentences alike. This is in line with the results of the omnibus ANOVA, which revealed a main effect of reversal and a reversal × anteriority interaction.

Event-related potentials (Control comparison)

The bottom half of Figure 2 shows the grand average ERPs and the topographic distribution of ERP effects in the control comparison. Results of the statistical analyses are presented in the middle column of Table 4.

As in Experiment 1, the cloze probability manipulation in the filler sentences elicited a clear N400 effect and a frontally distributed late positivity. Statistical analyses in the 350–450 ms time interval revealed a main effect of predictability and predictability × anteriority and predictability × anteriority × laterality interactions, due to the fact that expected words elicited a smaller N400 than unexpected words. Analyses in the 600–800 ms time interval revealed predictability × anteriority and predictability × anteriority × laterality interactions, due to the fact that unexpected words elicited a larger frontal positivity than expected words.

Discussion

In Experiment 2 we observed evidence that delaying the time between arguments and verbs by inserting a neutral adverbial phrase resulted in a significant argument role reversal N400 effect in the high-predictability conditions; this was not observed in the low-predictability conditions. Other aspects of the results remained the same as in Experiment 1. First, as in Experiment 1 argument role reversals were also readily detected and they elicited a clear P600 effect that did not differ between high- and low-predictability sentences. Second, the cloze probability manipulation in the filler sentences once again elicited a clear N400 effect followed by a late frontal positivity.

The most important finding in this experiment is the presence of an N400 effect in the high-predictability conditions but not in the low-predictability conditions. If these results are reliable, they would be incompatible with Semantic Illusion hypothesis, according to which comprehenders fail to detect implausibility in role-reversed sentences upon encountering the verb because they assign argument roles according to what is most plausible (e.g. Bornkessel-Schlesewsky & Schlesewsky, 2008; Hoeks et al., 2004; Kim & Osterhout, 2005; Kolk et al., 2003; Kuperberg, 2007). This is because the verb could always form a plausible thematic relation with the arguments in the experimental sentences regardless of its predictability, and thus if heuristics is at work then comprehenders should be able to compute a plausible interpretation and show no N400 effect in both conditions, just as in Experiment 1. It is possible that participants were sensitive to differences in how strongly (or weakly) constraining the sentence contexts were prior to the verb, but such early detection should only facilitate detection of the implausibility introduced by the verb and would not explain why the N400 effect emerged only when the verb was delayed. Further, these results would also be at odds with the hypothesis that comprehenders simply do not use argument role information in generating verb predictions. Instead, they would be in line with the Slow Prediction hypothesis and suggest that the N400 was insensitive to role-reversals in previous studies because comprehenders did not have sufficient time to incorporate information about the arguments' structural roles into their verb predictions.

Thus far our interpretation of these results is tentative. Even though the adverbial phrase was added in this experiment to increase the amount of time available for predictive computations and it did not affect either the cloze probability of the verbs or the constraint of the sentences, a within-participants comparison would allow a more direct test of the Slow Prediction hypothesis. Therefore, we manipulated the distance between the arguments and the verb within the same items and a single group of participants in Experiment 3.

Experiment 3

In this experiment we more directly tested the Slow Prediction hypothesis by combining aspects of Experiments 1 and 2 in a single design. As illustrated in Figure 3, in this experiment we varied the position of the adverbial time expression (e.g. "*shangxingqi*", last week) within the same set of materials. In the *short-distance conditions*, the time expression appeared at the beginning of the sentence and the SOA between the second argument and the verb was 600 ms, as in Experiment 1. Meanwhile, in the *long-distance conditions*, the time expression, along with the preposition "*zai*," appeared between the second argument and the target verb and the SOA between them became

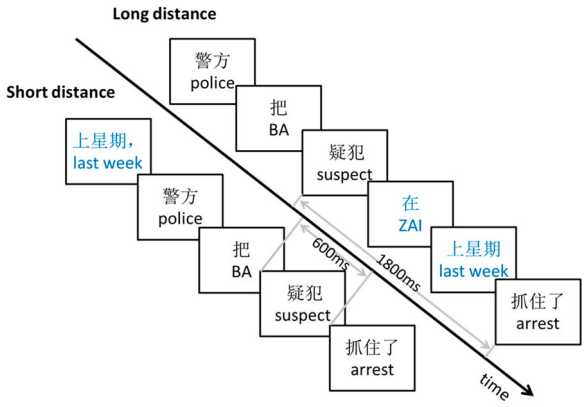

Figure 3. Distance manipulation in Experiment 3. Linear distance between the arguments and the verb was manipulated by placing a temporal phrase (e.g. last week) either at the beginning of the sentence (short-distance; 600 ms SOA between the second argument and the verb) or between the direct object and the verb (long-distance; 1800 ms SOA).

1800 ms, as in Experiment 2. Since the sentences in the short- and long-distance conditions were synonymous and differed only in the location of the temporal phrase, they should provide qualitatively similar contextual information for computing verb predictions. Under the Slow Prediction hypothesis, if argument role information can impact comprehenders' predictions for the verb within the extended time interval between the arguments and the verb, then the N400 should become sensitive to argument role reversals when the verb is further away from its arguments, i.e. in the long-distance conditions. Further, this experiment also provides another test of the Semantic Illusion hypothesis. Since the verb can form a plausible thematic relation with the arguments regardless of the linear distance between them, this hypothesis predicts no N400 effects to argument role reversals in either short- or long-distance conditions.

As the goal of Experiment 3 was to examine how time impacts the effect of predictive argument role information, we only included sentences in which argument role reversals had a clear impact on the verb's cloze probability, similar to the high-predictability conditions in Experiments 1 and 2. However, in order to reduce the likelihood of a floor effect on the N400, we tried to avoid strong lexical semantic associations between the verb and its arguments. To our knowledge no lexical association norms (or resources like Latent Semantic Analysis, Landauer & Dumais, 1997) are currently available in Mandarin Chinese, so we took high cloze probability values as a proxy for strong lexical associations and excluded sentences in which the verb had a very high cloze probability (> 70%) in the canonical condition.

Methods

Participants

Twenty-four students (22 female, mean age = 22 years, range 19–28 years) from South China Normal University participated in the current study. Informed consent was obtained from all participants. All were right-handed, native Mandarin Chinese speakers with normal or corrected-to-normal vision and were paid for their participation.

Materials

The stimuli consisted of 120 sets of sentences in the *ba*-construction in Mandarin Chinese. Argument order was manipulated in sentences in which the verb was closer vs. further away from the arguments, fully crossing distance (short vs. long distance) and reversal (canonical vs. role-reversed). Since the adverbial time expression (e.g. "*shangxingqi*", last week) appeared either at the start of the sentence (the *short-distance conditions*), or between the direct object and the verb along with the preposition "*zai*" (the *long-distance conditions*), the sentence contained the same content words before the verb in all four conditions, and the conditions differed only in the order of presentation. This effectively manipulated the linear distance, and thereby the amount of time, between the arguments and the verb without qualitatively changing the contextual information for computing verb predictions.

With the exception of items in which the verb had a cloze probability greater than 70% in the canonical condition, the stimuli largely overlapped with those in the high-predictability conditions in Experiments 1 and 2. As in Experiment 2, we gathered additional cloze probability norming data in order to verify that argument role reversals had a clear impact on the verb's cloze probability in short- and long-distance conditions alike and that placing the adverbial phrase after the arguments did not make the target verbs any more predictable. These were confirmed by the cloze data, which showed that there was a clear cloze probability difference between canonical and role-reversed sentences in both short-distance and long-distance conditions. In short-distance sentences, the target verb has an average target cloze probability of 37.4% (range = 16.7%–69%) in the canonical condition and always had 0% cloze in the role-reversed condition. In the long-distance sentences, average cloze probability was 31.0% (range = 0–86.7%) in the canonical condition and 0.4% (range = 0–13.3%) in the role-reversed condition.[8] Crucially, the difference in the verb's cloze probability between canonical and role-reversed sentences was smaller ($t(119) = 4.09$, $p < 0.01$), not larger, in the long-

distance conditions than in the short-distance conditions. In other words, if this cloze probability difference between the short- and long-distance conditions is meaningful, it actually creates a bias against our prediction that role reversals should have a bigger impact on the N400 in the long-distance conditions.[9]

Experimental sentences were distributed across four presentation lists, such that only one version of each item appeared in each list. Each list contained 120 experimental sentences (30 per condition) and the same set of 120 filler sentences used in Experiments 1 and 2. The sentences were presented in four blocks of 60 sentences each, and the order of the blocks was randomised across participants.

Procedure, EEG recording and ERP data analysis

The procedures for the experiment, EEG recording and data analysis were identical to those used in Experiments 1 and 2, with distance taking the place of predictability in the analyses. Although distance was manipulated within the same item sets in the current experiment, its effects are nonetheless not meaningful unless they interact with the effects of reversal. Therefore, only effects involving reversal are discussed in the main text. A total of 6.6% of experimental trials, roughly equally distributed across conditions (5% - 8.3%), were excluded from the averaging procedure due to artefacts.

Results

Plausibility judgments

Participants reliably judged canonical sentences to be plausible and the role-reversed sentences to be implausible with an overall accuracy of 89.5% (Short-distance, canonical = 93.4%; Short-distance, role-reversed = 87.5%; Long-distance, canonical = 89.9%; Long-distance, role-reversed = 86.7%).

Event-related potentials (Experimental comparisons)

The top half of Figure 4 shows the grand average ERPs at CPZ and the topographic distribution of the effects of argument role reversals at the target word in the 350–450 ms and 600–800 ms time intervals in the short-distance and long-distance conditions in Experiment 3. Results of the posterior electrode cluster analyses and the omnibus ANOVA are presented in Table 6 and the rightmost column of Table 3 respectively. Grand average ERPs across all 30 electrode sites are presented in Supplementary Figures 5 and 6.

In Experiment 3, an N400 effect of role-reversal was elicited in the long-distance conditions but not in the short-distance conditions. In the 350–450 ms interval, the posterior electrode cluster analyses showed a significant N400 effect in the long-distance conditions but not in the short-distance conditions, as predicted by the hypothesis that argument role information has a delayed impact on comprehenders' verb predictions. This is in line with the results of the omnibus ANOVA, which revealed a reversal × distance × laterality interaction, reflecting a larger negativity over central and right posterior electrodes in the long-distance condition than in the short-distance condition.

In the 600–800 ms interval, the posterior cluster analyses showed that role reversal elicited a significant P600 effect in long- and short-distance conditions alike. This is consistent with the omnibus ANOVA, which revealed a main effect of reversal, a reversal × anteriority interaction, but no interactions involving distance and reversal.

Event-related potentials (Control comparison)

Grand average ERPs and the topographic distribution of ERP effects in the control comparison are presented on the bottom half of Figure 4. Results of the statistical analyses are presented in the rightmost column of Table 4.

As in Experiments 1 and 2, the standard cloze probability manipulation in the filler sentences elicited a clear N400 effect and a frontally distributed late positivity. Statistical analyses in the 350–450 ms time interval revealed a main effect of predictability and predictability × anteriority and predictability × anteriority × laterality interactions, showing that expected words elicited a smaller central posterior negativity (N400) than unexpected words. Analyses in the 600–800 ms time interval revealed predictability × anteriority and predictability × anteriority × laterality interactions, showing that unexpected words elicited a larger frontal positivity than expected words.

Discussion

Figure 5 presents a summary of the main results of Experiments 1–3. The results of Experiment 3 demonstrated that argument role reversals impacted the amplitude of the N400 at the target verb when it was further away from the arguments (in the long-distance conditions) but not when it immediately followed the arguments (in the short-distance conditions), replicating key findings of Experiments 1 and 2. Other aspects of the results were also consistent with those of Experiments 1 and 2. First, argument role reversals were readily detected in both short- and long-distance conditions and elicited a robust P600 effect in both short- and long-distance conditions. Second, the cloze probability manipulation in the filler sentences once again elicited a clear N400 effect followed by a late frontal positivity.

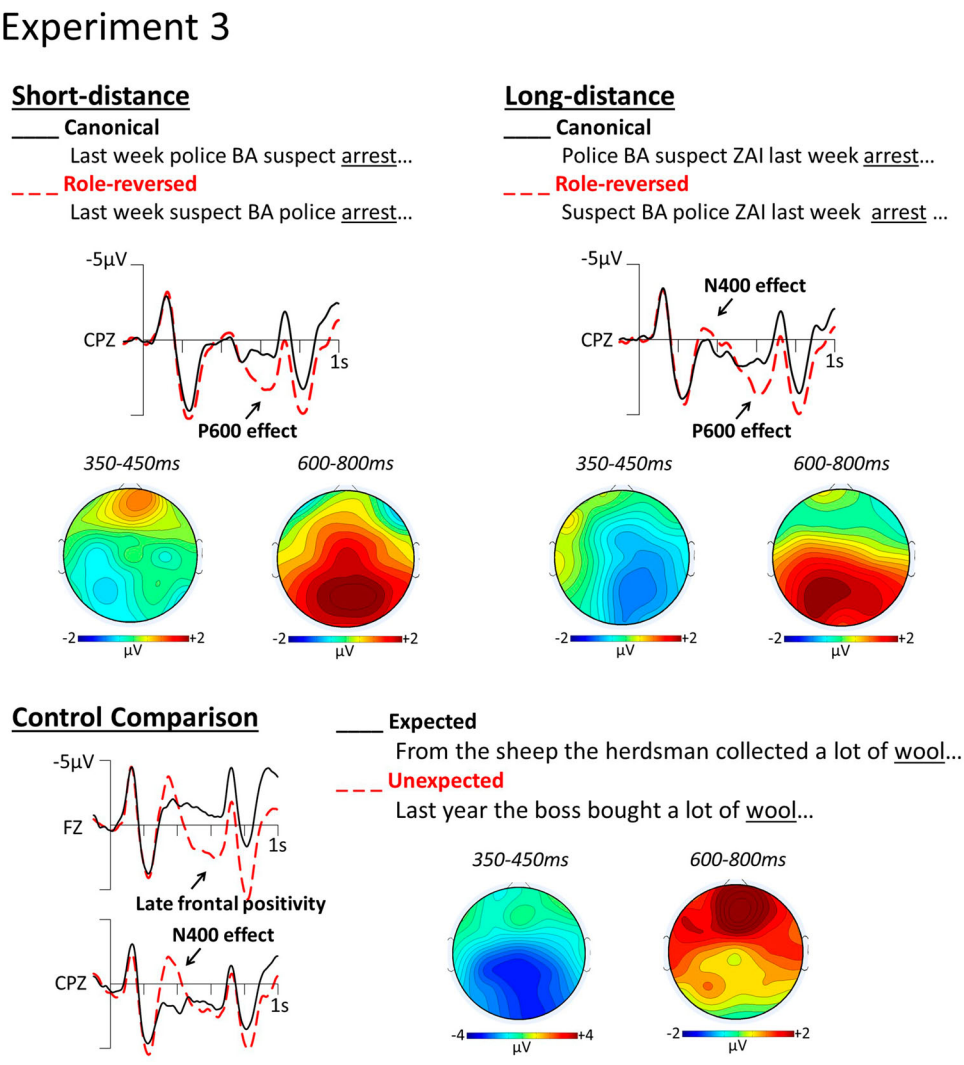

Figure 4. *Top*: Grand average ERPs at centro-posterior electrode CPZ and topographic distribution of ERP effects in the 350–450 ms and 600–800 ms intervals (role-reversed minus canonical) in the short-distance (left) and long-distance (right) conditions in Experiment 3. Argument role reversals elicited a P600 effect in the short-distance conditions, and an N400 effect followed by a P600 effect in the long-distance conditions. The N400's sensitivity to the target verb's offline predictability critically depended on the time elapsed between the arguments and the verb. When the verb appears immediately after the arguments, expected and unexpected verbs elicit no N400 difference. When an adverbial phrase appeared between the arguments and the verb, the N400 is reduced for the expected verb compared to the unexpected verb. *Bottom*: Grand average ERPs at frontal electrode FZ and at centro-posterior electrode CPZ and topographic distribution of ERP effects in the 350–450 ms and 600–800 ms intervals (expected minus unexpected) in the control comparison in Experiment 3. A cloze probability manipulation in the filler sentences elicited a clear N400 effect followed by a frontally distributed late positivity.

Table 6. Results of the posterior electrode cluster analyses in Experiment 3: *t* value, standardised effect size (Pearson's *r*) and 95% confidence interval for the contrast between the canonical and role-reversed conditions in each time interval.

Experiment 3	*df*	*t*	Pearson's *r*	95% CI
Short distance				
350–450 ms	23	<1	−0.14	[−0.51, 0.31]
600–800 ms	23	3.6**	0.60	[0.21, 0.80]
Long distance				
350–450 ms	23	2.24*	−0.42	[−0.68, −0.005]
600–800 ms	23	4.21**	0.66	[0.32, 0.82]

**$p < .01$; *$p < .05$.

The N400's sensitivity to argument role reversals in the long-distance conditions stands in sharp contrast with its insensitivity in the short-distance conditions and in many previous studies. Since the sentences in the short- and long-distance conditions were synonymous and differed only by the location of the temporal phrase, the change in the N400's sensitivity is not attributable to changes in the content of the context per se. Further, since the distance manipulation did not affect the cloze probability difference between the canonical

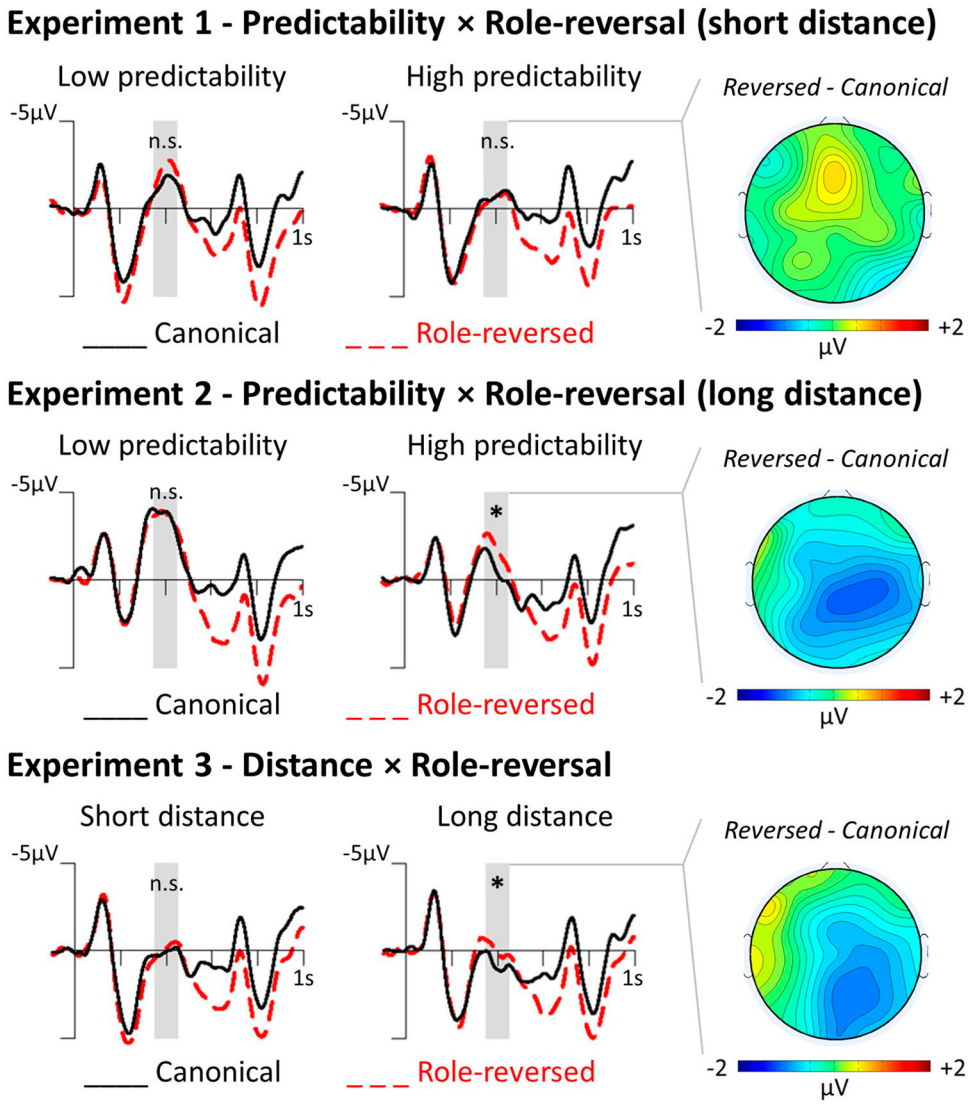

Figure 5. The left and centre columns show average ERPs at centro-posterior electrode CPZ to the target verb (350–450 ms highlighted) in Experiments 1–3. The right column shows, from top to bottom, the topographic distribution of the difference between role-reversed and canonical conditions during the N400 time interval (350–450 ms) in the high-predictability conditions in Experiments 1 and 2, and in the long-distance conditions in Experiment 3. Argument role reversals elicited a P600 effect in all cases, likely reflecting the detection of implausibility. Meanwhile, argument role reversals elicited a significant N400 effect at the verb only when they resulted in a clear difference in the verb's offline predictability and when the target verb was further away from the arguments.

and role-reversed sentences, the observed difference between the short- and long-distance conditions cannot be attributed to a change in the target verb's cloze probability either. Instead, it is likely due to the additional time elapsed between the second argument and the verb (600 ms vs. 1800 ms SOA) that was introduced by the distance manipulation. We take the presence of an N400 effect in the long-distance conditions to indicate that access to the target verb's meaning was facilitated in the canonical sentences relative to the role-reversed sentences when additional time elapsed between the arguments and the verb. Further, we argue that the fact that the N400 effect only emerged in the long-distance conditions constitutes supporting evidence for the Slow Prediction Hypothesis, according to which information about the arguments' roles has a delayed impact on comprehenders' verb predictions.

General discussion

In the current study, we investigated the possibility that even salient and unambiguous information about preverbal arguments' structural roles may have a delayed impact on online verb predictions (the Slow Prediction hypothesis). We manipulated the linear distance between the arguments and the verb to examine how it may affect comprehenders' response to the verb in

canonical vs. role-reversed sentences in a series of ERP experiments.

Across all of the current experiments, the implausibility created by reversal of pre-verbal arguments was readily detected and it consistently elicited a robust P600 effect at the verb. These results are in line with previous findings (Chow & Phillips, 2013; Chow, Smith, et al., 2016; Hoeks et al., 2004; Kolk et al., 2003; van Herten et al., 2006, 2005), and show that comprehenders reliably use argument role information to interpret thematic relations upon encountering the verb.

Meanwhile, argument role reversals elicited an N400 effect only when two conditions were met: argument reversal had a strong impact on the verb's cloze probability and the verb was further away from the arguments. In Experiment 1 we extended previous findings by showing that argument role reversals did not modulate the N400 at the verb even when they resulted in a large cloze probability difference between conditions. In Experiments 2 and 3 we found that argument role reversals elicited an N400 effect when (i) there was a clear cloze probability differences between conditions and when (ii) the linear distance and, effectively, the time interval between the arguments and the verb was widened. Note, however, that the N400 did not simply become sensitive to argument role reversals when the verb was further away from its arguments: argument role reversals did not elicit an N400 effect in the low-predictability sentences even with the added distance. This strongly suggests that the N400 effect that emerged in the long-distance conditions reflected sensitivity to the verb's offline predictability. These distance effects are also in line with findings in a recent study in Japanese that directly manipulated SOA in simple SV and OV sentences (Momma, Sakai, & Phillips, 2015). Taken together, these findings have two main implications.

Firstly, the present findings are incompatible with the Semantic Illusion hypothesis, according to which comprehenders are initially blind to the implausibility in role-reversed sentences when they first encounter the verb (e.g. Bornkessel-Schlesewsky & Schlesewsky, 2008; Hoeks et al., 2004; Kim & Osterhout, 2005; Kolk et al., 2003; Kuperberg, 2007). This hypothesis, which assumes the integration view of the N400, would expect the N400 to be insensitive to argument role reversals regardless of the distance between the verb and the arguments because they could form a plausible thematic relation in all cases. However, the observation of an N400 effect in the high-predictability conditions in Experiment 2 and in the long-distance condition in Experiment 3 shows that argument role reversals can elicit an N400 effect even when the verb can form a plausible thematic relation with the arguments.

On the other hand, these findings are compatible with the prediction view of the N400 and provide support for the Slow Prediction hypothesis, according to which argument role information cannot impact comprehenders' verb predictions immediately. The effect of distance suggests that argument role information can inform comprehenders' verb predictions, but that this requires sufficient time between the arguments and the verb. Further, the observation that the N400 became sensitive to argument role reversals only when there was a clear difference in the verb's cloze probability between conditions suggests that it is changes in comprehenders' predictions that underlies the N400's re-emerged sensitivity.

The present study extends previous eye-tracking findings by Kukona et al. (2011), who found that listeners' anticipatory eye-movements were sensitive to filled vs. unfilled thematic roles in passive sentences but not in active sentences. First, whereas Kukona and colleagues compared passive and active sentences that differed in multiple ways, in the present study we directly manipulated the distance between the arguments and the verb in active sentences by only changing the position of the temporal adverbial. This allowed us to keep the rest of the sentence constant and examine the effect of timing more directly. Meanwhile, instead of presenting only a small set of objects on the screen and tracking listeners' looks to these objects as they process the linguistic stimuli, in the present study we measured comprehenders' ERP response as they read. This allowed us to examine comprehenders' response to the linguistic stimuli without artificially restricting their options. Further, while the extent to which listeners' anticipatory fixations are reflective of comprehenders' predictions vs. other cognitive processes remains debatable, there is much evidence that the N400 ERP response is modulated by the ease of accessing a word's meaning from long-term memory (Kutas & Federmeier, 2000; Lau et al., 2008) and its well-attested sensitivity to a word's predictability makes it an excellent measure of the extent to which a word's meaning is pre-activated.

More broadly, our observation that comprehenders' verb predictions are initially not sensitive to information about the arguments' structural roles calls into question a simplifying assumption that is common among existing models of prediction, namely, that all relevant information can impact prediction immediately (e.g. Altmann & Mirković, 2009; Demberg & Keller, 2008a; Elman, 1990; Hale, 2001; Levy, 2008; Smith & Levy, 2013). Instead, they provide empirical support for Laszlo and Federmeier's (2009) proposal that "quantitative shifts in the timing of processing can potentially lead to qualitative

differences in what particular facets of semantics come to be linked up with a given input" (p. 32) and highlight the importance of timing considerations in the study of prediction (cf. Dambacher et al., 2012; Ito, Corley, et al., 2016; Wlotko & Federmeier, 2015). Further, the widened time interval in Experiments 2 and 3 (1800 ms between the onset of the second argument and the verb) places an upper limit on the time required for argument role information to impact verb predictions. By isolating the time course of a specific component of predictive processing, these findings open up a host of questions about how different sources of information are integrated to feed linguistic prediction in real time.

Note, however, that the present proposal is distinct from "good-enough" models of language comprehension (e.g. Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, Bailey, & Ferraro, 2002), or any model which posits that detailed syntactic information may be ignored during comprehension (e.g. Townsend & Bever, 2001). As we argued above, comprehenders' accurate plausibility judgments and the presence of a robust P600 effect across all experiments suggest that they reliably use argument role information for interpreting the sentences in real time. Meanwhile, the N400 results suggest that it is specifically the comprehenders' predictions that are not immediately sensitive to argument role information. Under this account, comprehenders immediately and reliably use argument role information to interpret thematic relations when the verb appears, but they may not use this information quickly enough to predict the upcoming verb before it appears in the input.

The role of lexical semantic association

Just as listeners' eye-movements are sensitive to semantic relationships between the linguistic stimuli and the visual objects (e.g. Huettig & Altmann, 2005; Yee & Sedivy, 2006), the amplitude of the N400 response to a word is reduced when it is preceded by semantically related words (for a review see Ledoux, Camblin, Swaab, & Gordon, 2006). Therefore, it is important to consider how lexical semantic association between the arguments and the verb might have contributed to the results observed here. In particular, one might attribute the absence of N400 effects in the short-distance conditions and in previous studies to strong lexical semantic associations between the arguments and verbs (Brouwer et al., 2012). If such associations had rendered the N400 at the verb at floor, then we would not be able to observe any N400 effect at the verb, even if argument role information did have an immediate impact on verb predictions. According to this explanation, the N400's sensitivity could have re-emerged in the long-distance conditions not because of the amount of time available for prediction, but because the effect of lexical association from the arguments decayed during the adverbial phrase (e.g. Ratcliff, Hockley, & McKoon, 1985). This is an interesting alternative possibility that warrants further investigation in future work. However, we note that the average N400 at the target words in the experimental sentences was consistently much larger than that elicited by the expected words in the filler sentences across all three experiments, and for this reason we consider it unlikely that the N400 response to the verb in the short-distance conditions was at floor.[10]

Late frontal positivity and the P600

Although a late positivity was observed in experimental and control comparisons across Experiments 1–3, these effects had clearly different topographic distributions. In the experimental sentences, where the role-reversal manipulation could affect both predictability and plausibility, target verbs elicited a larger late positivity in the role-reversed condition than in the canonical condition, regardless of the predictability of the verb (Experiments 1 and 2) or the distance between the arguments and the verb (Experiment 3). In all cases this late positivity was most pronounced at parietal scalp sites, which is typical of a P600 effect. Meanwhile in the filler sentences, where predictability was modulated but plausibility was held constant, the late positivity elicited by unexpected words was frontally distributed across all three experiments. The different topographies of these effects suggest that they might have arisen from different brain regions and might reflect different functional processes.

While a proposal by Kuperberg (2013) has linked these late positivities to different kinds of prediction error, a meta-analysis by van Petten and Luka (2012) suggests that only the frontal positivity is associated with prediction. They observed that late positivities to semantically incongruous words tend to be posteriorly distributed and that late positivities to unexpected but nonetheless semantically congruous words are often frontally distributed (van Petten & Luka, 2012). In the current study, argument role reversals resulted in implausible experimental sentences, whereas all of the sentences in the control comparison were plausible. The observation of a P600 effect in the experimental comparisons but not in the control comparisons is in line with van Petten and Luka's (2012) observation and suggests that the P600 is not modulated by a word's predictability per se, but by the plausibility or congruence of the sentence. Therefore, the P600 effect likely reflects processes that

are triggered in case of comprehension failures (e.g. upon detecting implausibility or grammatical anomalies in the input), and these processes might involve re-analyses (e.g. Friederici, 1995), context updating (Coulson et al., 1998), and/or error corrections in a noisy channel model (e.g. Gibson, Bergen, & Piantadosi, 2013; Kim & Sikos, 2011).

Meanwhile, results in the control comparisons are consistent with previous reports that unexpected but congruent words elicited a larger N400 and a frontal positivity relative to strongly expected words (e.g. Delong, Urbach, Groppe, & Kutas, 2011; Federmeier et al., 2007; Thornhill & van Petten, 2012). Previous research has proposed that the frontal positivity reflects the suppression or inhibition of a prediction for a different word or concept (Federmeier et al., 2007; Kutas, 1993) and/or learning signal from which to update future predictions (Federmeier, Kutas, & Schul, 2010). However, existing results have remained mixed with regard to whether the positivity becomes larger when a strong prediction is violated (Federmeier et al., 2007), or if the amplitude of the frontal positivity is reduced when a strongly expected word is encountered (Thornhill & van Petten, 2012). Therefore, future research will be needed to determine exactly how this component is related to predictive processes.

Other cases of apparent "prediction failure"

The present study demonstrates that even prominent and unambiguous contextual information such as word order may fail to immediately impact predictions even in healthy young adult native speakers. This reinforces an important but often overlooked distinction between a word's offline predictability (cloze probability) and its online predictability (Chow, Smith, et al., 2016). Since offline predictability measures obtained in language corpora and offline cloze tasks do not capture the temporal aspects of linguistic prediction, models of language comprehension that are based on offline predictability measures would fail to capture cases in which some contextual information has a non-immediate impact on comprehenders' online predictions.

Further, it highlights the fact that a prediction can facilitate processing only if it precedes the predicted event, and as such careful considerations about timing are crucial for differentiating between genuine failures to engage predictive mechanisms and cases in which specific sources of contextual information fail to impact predictions quickly enough to facilitate bottom-up processing. For example, recent studies have reported that predictive processing effects are absent under certain circumstances (e.g. Chwilla, Virgillito, & Vissers, 2011) and in different speaker groups such as children with low vocabulary scores (e.g. Borovsky, Elman, & Fernald, 2012; Mani & Huettig, 2012), older adults (DeLong, Groppe, Urbach, & Kutas, 2012; Federmeier & Kutas, 2005; Federmeier, McLennan, De Ochoa, & Kutas, 2002; Wlotko, Federmeier, & Kutas, 2012), second language learners (e.g. Grüter, Lew-Williams, & Fernald, 2012; Kaan, 2014; Martin et al., 2013), illiterate adults (Mishra, Singh, Pandey, & Huettig, 2012), and patients with schizophrenia (e.g. Ford & Mathalon, 2012; Kuperberg, 2010). While such findings may suggest that certain speaker groups cannot (or do not) engage predictive processing, it is also possible that these speakers do in fact anticipate upcoming inputs during comprehension, but that some of the computations involved are still incomplete when the relevant input arises.

To take a more vivid analogy, imagine throwing a ball to a child. He might fail to catch it because he lacks the ability to predict where it is going, but it could also be because he simply cannot compute those predictions quickly enough in order to move to the predicted position before the ball gets there. Therefore, to examine potential causes for apparent prediction failures we must isolate sub-components of predictive computations and examine how different sources of contextual information impact comprehenders' prediction as the input unfolds in real time. For example, future research can examine whether second language speakers are more or less likely to show predictive effects depending on the type of contextual information involved (e.g. grammatical vs. Pragmatic cues) and the amount of the time they are given to process it (e.g. Ito, Martin, & Nieuwland, 2016).

Towards an explicit model of predictive processing

We believe that the current experimental paradigm can be applied more broadly to examine the time course of predictive computations. This paradigm has three key components. First, it requires isolating the source of contextual information under investigation. Second, it manipulates the timing at which this piece of information becomes available relative to a target word. Third, it requires a dependent measure that is modulated by the output of predictive processing. The time course with which the contextual information of interest impacts comprehenders' prediction can then be inferred by examining this dependent measure's sensitivity at different time points. By controlling the strength and timing of specific predictive cues, this paradigm can be applied more broadly to examine the time course with which different information sources impact predictive

processing and how it might be affected in different populations.

For example, in a recent study in English, we expanded this paradigm to examine the possibility that lexical semantic information about the arguments can impact comprehenders' verb predictions more quickly than information about their structural roles (Chow, Smith, et al., 2016). We manipulated the verb's cloze probability by either (i) substituting one of the preverbal arguments (e.g. "The superintendent overheard which tenant/realtor the landlord had evicted … ") or (ii) reversing the order of the arguments (e.g. "The restaurant owner forgot which customer/waitress the waitress/customer had served … "). Crucially, even though these manipulations had the same impact on the verb's cloze probability, an N400 effect was observed in the argument substitution condition only. This suggests the arguments' lexical meaning may have a more immediate impact on verb predictions than their structural roles, and, more generally, that different sources of information may impact linguistic predictions on different time scales.

Why might information about argument roles have a delayed impact on verb predictions compared to lexical semantic information about the arguments? Here we briefly consider three different ways in which the impact of argument role information on the process of retrieving likely events from memory may be delayed. For a more detailed discussion, see Chow, Momma, et al. (2016).

One possibility is that the delay originates at the start of the memory retrieval process, that is, when retrieval cues are issued. In languages such as Mandarin Chinese and English, arguments' structural roles are determined in large part by word order, and comprehenders may be able to access the arguments' lexical representation (which is independent of word order) before they can determine the arguments' structural roles. In turn, a delay could arise if comprehenders start to retrieve from memory events that involve those arguments as participants even before their structural roles are identified. Under this view, even a slight delay in when the relevant retrieval cues are issued could have a cascading effect through the retrieval process and result in a much bigger delay in their impact on comprehenders' verb predictions. However, this account might seem less plausible in light of recent studies in Japanese using case-marking as a cue to argument role (Momma et al., 2015); although it seems likely that case information could be processed rapidly, they similarly report evidence that argument role information cannot impact comprehenders' verb prediction within 800 ms from argument onset.

Another possibility is that the delay may lie in the dynamics of the competition among predictive cues (Kuperberg, 2016) and/or partially matching candidates (Lewis, Vasishth, & Van Dyke, 2006). Under one version of this type of account, argument role information may impact verb predictions more quickly when it is more prominent or strongly primed by the preceding context. For example, if a preceding question asked "Who was doing what?", focusing the argument role information, readers might be faster to process this information and incorporate it into predictions in the subsequent sentence.

The third possibility is that argument role information cannot be used to query event knowledge directly, due to the way in which event knowledge is encoded in memory. Under this view, there may be information in the linguistic input that constrains which events are likely to be discussed, but that cannot be directly used as a cue for retrieving those events from memory. For example, while arguments are associated with grammatical roles such as subject and object in the linguistic input, event representations naturally do not encode event participants in those terms. Therefore, at a minimum comprehenders would have to use an argument and its grammatical role to identify its likely thematic role (e.g. animate subjects are relatively likely to be agents) before they can query events that take this argument in a specific role. Furthermore, it remains controversial whether sentence or verb meanings make reference to abstract thematic roles such as agent and patient, or equally whether such roles are fundamental primitives of non-linguistic event representations (see Williams, 2015 for discussion). As such, when arguments are encountered prior to the verb, a more indirect mechanism is likely required to use grammatical or thematic role information to retrieve events from memory.

In Chow, Momma, et al. (2016), we discussed the possibility that using argument role information for verb prediction might involve an indirect mechanism, in which arguments (or event participants) are initially used to retrieve events (and the verbs associated with those events), irrespective of their roles (e.g. both serving and tipping events are likely to involve customers and waitresses as participants), and that the set of candidate verbs is then evaluated to pick out those that match the specific roles of the arguments (e.g. a tipping event is likely given the customer-as-agent and waitress-as-patient cues but a serving event is not). Further research will be needed to investigate how event knowledge is encoded in long-term memory and the way(s) in which such stored representations may be accessed during language comprehension.

Conclusion

The current study investigates the possibility that even salient information about preverbal arguments' roles may have a delayed impact on online verb predictions. By investigating how the timing of argument role information impacts comprehenders' online verb predictions, we show that argument role information does contribute to verb prediction, but the predictive computations involved may be relatively slow. These results highlight the importance of timing considerations in studying how diverse contextual information contributes to predictions.

Notes

1. Federmeier and Kutas (1999) pointed out that the predictability of a *word*, as assessed by cloze probability, is not always identical to the predictability of the *meaning* (or semantic features) of that word. Therefore, cloze probability may be an imperfect proxy if the N400 amplitude is modulated by the ease of accessing the *meaning* of a word. We return to this point in discussing the logic of the current study.
2. Earlier studies have reported that the N400 was insensitive to negation in simple sentences (e.g. "A robin is not a tree/bird"; Fischler, Bloom, Childers, Roucos, & Perry, 1983; Katayama, Miyata, & Yagi, 1987; Kounios & Holcomb, 1992; Lüdtke, Friedrich, De Filippis, & Kaup, 2008), but more recent work has demonstrated that the N400 is sensitive to negation as long as it is pragmatically licensed (Nieuwland & Kuperberg, 2008).
3. These accounts typically posit that such insensitivity to implausibility is temporary, as participants' end-of-trial plausibility judgements are highly accurate across studies.
4. In Experiment 1, average sentence constraint in the canonical condition was 64% in the high-predictability sentences and 22% in the low-predictability sentences; in the role-reversed condition, average sentence constraint was 25% and 22% in high- and low-predictability sentences respectively. The target verb was the most predictable word in the canonical condition in all of the high predictability sentences and only some of the low predictability sentences.
5. Note that this result does not indicate that comprehenders computed predictions more quickly in the filler sentences than in the experimental sentences. As in previous studies on the effects of cloze probability, the cloze probability manipulation in the filler sentences did not allow precise estimation of the delay between predictive context and predicted target because the same target words were presented in completely different sentence contexts across conditions.
6. Since these sentences were adapted from those used in Experiment 1, some target words in the low predictability sentences (which had very low cloze probabilities in Experiment 1) may have a zero cloze probability simply due to sampling error.
7. Sentence constraint were also highly similar between items in Experiments 1 and 2. In Experiment 2, average sentence constraint (cloze probability of the most likely continuation) in the canonical condition was 58% in the high-predictability sentences and 21% in the low-predictability sentences; in the role-reversed condition, average sentence constraint was 26% and 22% in high-predictability and low-predictability sentences respectively.
8. Since the materials were developed by first selecting the short-distance sentences with a suitable cloze probability profile and then adapting those sentences to create the long-distance versions, in some cases the target word had a lower cloze probability in the long-distance condition simply because a synonym (or a closely related word) was more frequent in the cloze responses in the long-distance sentences.
9. As in Experiments 1 and 2, sentence constraint was not affected by the distance manipulation in Experiment 3. In Experiment 3, average sentence constraint in the canonical condition was 37% and 35% in the short- and long-distance sentences respectively; in the role-reversed condition, average sentence constraint was 27% in both short- and long-distance sentences. The target word was the most predictable word in the canonical condition in the majority of the items.
10. Since the target words in the filler sentences were not matched in any way with those in the experimental sentences, such a comparison may be suggestive but must be interpreted with caution.

Acknowledgements

We thank Nan Li, Lan Chen, Jie Li, Qiong Sun, Wenjia Zhang, Luodi Yu, Jian Wang, Zhizhou Deng, Dongxia Sun, Xiayan Zhu, Yangling Cui and Gangyi Feng for invaluable help in material creation and data collection.

Disclosure statement

No potential conflict of interest was reported by the authors .

Funding

This work was supported in part by a National Science Foundation grant (BCS-0848554) to Colin Phillips, and by a Key Project of National Social Science Foundation of China (15AZD048) and a Key Project of National Natural Science Foundation of Guangdong Province, China (2014A030311016) to Suiping Wang.

ORCID

Wing-Yee Chow http://orcid.org/0000-0002-1734-485X

References

Altmann, G. T. M. (1999). Thematic role assignment in context. *Journal of Memory and Language*, *41*(1), 124–145. doi:10.1006/jmla.1999.2640

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. doi:10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33* (4), 583–609. doi:10.1111/j.1551-6709.2009.01022.x

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, *51*(1), 23–54. doi:10.1016/j.jml.2004.03.001

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. doi:10.1016/j.tics.2007.05.005

Bar, M. (Ed.). (2011). *Predictions in the brain: Using our past to generate a future*. New York: Oxford University Press.

Bendixen, A., Schroger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, *29*(26), 8447–8451. doi:10.1523/JNEUROSCI.1493-09.2009

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, *63*(4), 489–505. doi:10.1016/j.jml.2010.08.004

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., … Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, *117* (3), 133–152. doi:10.1016/j.bandl.2010.09.010

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. doi:10.1016/j.brainresrev.2008.05.003

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. doi:10.1016/j.jecp.2012.01.005

Brouwer, H., Fitz, H., & Hoeks, J. C. J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. doi:10.1016/j.brainres.2012.01.055

Brown, C. M, & Hagoort, P. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44.

Brusini, P., Brun, M., Brunet, I., & Christophe, A. (2015). Listeners exploit syntactic structure on-line to restrict their lexical search to a subclass of verbs. *Frontiers in Psychology*, *6* (December), 1841. doi:10.3389/fpsyg.2015.01841

Chow, W.-Y., Momma, S., Smith, C., Lau, E., & Phillips, C. (2016). Prediction as memory retrieval : Timing and mechanisms. *Language, Cognition and Neuroscience*, *31*(5), 617–627. doi:10.1080/23273798.2016.1160135

Chow, W.-Y., & Phillips, C. (2013). No semantic illusions in the semantic P600 phenomenon: ERP evidence from mandarin Chinese. *Brain Research*, *1506*, 76–93. doi:10.1016/j.brainres.2013.02.016

Chow, W.-Y., Smith, C., Lau, E., & Phillips, C. (2016). A "bag-of-arguments" mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, *31*(5), 577–596. doi:10.1080/23273798.2015.1066832

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

Chwilla, D. J., Virgillito, D., & Vissers, C. T. W. M. (2011). The relationship of language and emotion: N400 support for an embodied view of language comprehension. *Journal of Cognitive Neuroscience*, *23*(9), 2400–2414. doi:10.1162/jocn.2010.21578

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21–58. doi:10.1080/016909698386582

Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: Event-related potentials during sentence reading. *Neuropsychologia*, *50*(8), 1852–1870. doi:10.1016/j.neuropsychologia.2012.04.011

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*(1), 89–103. doi:10.1016/j.brainres.2006.02.010

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, *121*(3), 226–239. doi:10.1016/j.bandl.2012.02.006

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*(1), 150–162. doi:10.1016/j.neuropsychologia.2014.06.016

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*(9), 1203–1207. doi:10.1111/j.1469-8986.2011.01199.x

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

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. doi:10.1016/j.cognition.2008.07.008

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkanen, L. (2010). Early occipital sensitivity to syntactic category Is based on form typicality. *Psychological Science*, *21*(5), 629–634. doi:10.1177/0956797610367751

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition*, *110*(3), 293–321. doi:10.1016/j.cognition.2008.09.008

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. doi:10.1207/s15516709cog1402_1

Farmer, T. A., Yan, S., Bicknell, K., & Tanenhaus, M. K. (2015). Form-To-Expectation matching effects on first-pass Eye movement measures during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 958–976. doi:10.1037/xhp0000054

Federmeier, K. D., & Kutas, M. (1999). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*(3), 373–392. doi:10.1016/S0926-6410(99)00036-1

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, *42*(2), 133–141. doi:10.1111/j.1469-8986.2005.00274.x

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during

language comprehension. *Brain and Language*, *115*(3), 149–161. doi:10.1016/j.bandl.2010.07.006

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39* (2), 133–146. doi:10.1017/S0048577202001373

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*(1), 75–84. doi:10.1016/j.brainres.2006.06.101

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*(4), 516–547. doi:10.1006/jmla.2000.2728

Fischler, I, Bloom, P. A, Childers, D. G, Roucos, S. E, & Perry, N. W. J. (1983). Brain Potentials Related to Stages of Sentence Verification. *Psychophysiology*, *20*(4), 400–409.

Ford, J. M., & Mathalon, D. H. (2012). Anticipating the future: Automatic prediction failures in schizophrenia. *International Journal of Psychophysiology*, *83*(2), 232–239. doi:10.1016/j.ijpsycho.2011.09.004

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*. doi:10.1006/brln.1995.1048

Gerlanc, D., & Kirby, K. N. (2013). *bootES*. Retrieved from http://cran.r-project.org/web/packages/bootES/index.html

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America*, *110* (20), 8051–8056. doi:10.1073/pnas.1216438110

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112. doi:10.1007/BF02289823

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215. doi:10.1177/0267658312437990

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*, 660–676.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483.

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

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis : early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Proceedings of NAACL 2001*, *2*, 159–166. doi:10.3115/1073336.1073357

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, *111*(2), 151–167. doi:10.1016/j.cognition.2009.01.009

Hawkins, J., & Blakeslee, S. (2004). *On intelligence*. New York: Times Books.

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*(1), 59–73. doi:10.1016/j.cogbrainres.2003.10.022

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32. doi:10.1016/j.cognition.2004.10.003

Ilkin, Z., & Sturt, P. (2011). Active prediction of syntactic information during sentence processing. *Dialogue and Discourse*, *2*(1), 35–58. doi:10.5087/dad.2011.103

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

Ito, A., Martin, A. E., & Nieuwland, M. S. (2016). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 635–652. doi:10.1037/xlm0000315

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. doi:10.1075/lah-4-2.05kaa

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

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. doi:10.1016/S0749-596X(03)00023-8

Katayama, J, Miyata, Y, & Yagi, A. (1987). Sentence verification and event-related brain potentials. *Biological Psychology*, *25* (2), 173–185. doi:10.1016/0301-0511(87)90036-6

Kim, A., & Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience*, *24*(5), 1104–1112. doi:10.1162/jocn_a_00148

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. doi:10.1016/j.jml.2004.10.002

Kim, A., & Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language*, *118*(1-2), 15–22. doi:10.1016/j.bandl.2011.03.002

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. doi:10.1016/j.cognition.2004.03.002

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85* (1), 1–36. doi:10.1016/S0093-934X(02)00548-5

Kos, M., Vosse, T., van den Brink, D., & Hagoort, P. (2010). About edible restaurants: Conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Psychology*, *1*(DEC), 1–11. doi:10.3389/fpsyg.2010.00222

Kounios, J, & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain

potentials and reaction times. *Journal of Experimental Psychology: General*, *121*(4), 459–479.

Kukona, A., Fang, S. Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23–42. doi:10.1016/j.cognition.2010.12.002

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*(1), 23–49. doi:10.1016/j.brainres.2006.12.063

Kuperberg, G. R. (2010). Language in schizophrenia part 1: An introduction. *Linguistics and Language Compass*, *4*(8), 576–589. doi:10.1111/j.1749-818X.2010.00216.x

Kuperberg, G. R. (2013). The proactive comprehender: What event-related potentials tell us about the dynamics of reading comprehension. Gina. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the behavioral, neurobiological, and genetic components of reading comprehension* (pp. 176–119). Baltimore: Paul Brookes Publishing.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2015.1130233

Kuperberg, G. R., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117–129. doi:10.1016/S0926-6410(03)00086-7

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*(4), 533–572. doi:10.1080/01690969308407587

Kutas, M, DeLong, K. A, & Smith, N. J. (2011). A Look around at What Lies Ahead: Prediction and Predictability in Language Processing. In *Predictions in the Brain: Using Our Past to Generate a Future* (pp. 190–207). New York: Oxford University Press doi:10.1093/acprof:oso/9780195395518.003.0065

Kutas, M., & Federmeier, K. D. (2000). Electropsysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, *4*(12), 463–470. doi:10.1016/S1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(August), 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*(4427), 203–205. doi:10.1126/science.7350657

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

Landauer, T. K, & Dumais, S. T. (1997). A solution to Plato's problem : The Latent Semantic Analysis Theory of Acquisition. *Induction , and Representation of Knowledge. Psychological Review*, *104*(2), 211–240. doi:10.1037/0033-295X.104.2.211

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 and Language*, *61*(3), 326–338. doi:10.1016/j.jml.2009.06.004

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. doi:10.1038/Nrn2532

Lau, E, Stroud, C, Plesch, S, & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, *98*(1), 74–88. doi:10.1016/j.bandl.2006.02.003

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. *Behavioral and Cognitive Neuroscience Reviews*, *5*(3), 107–127. doi:10.1177/1534582306289573

Levy, R. P. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. doi:10.1016/j.cognition.2007.05.006

Levy, R. P., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, *122*(1), 12–36. doi:10.1016/j.cognition.2011.07.012

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. doi:10.1016/j.tics.2006.08.007

Llinás, R. (2002). *I of the vortex: From neurons to self*. Cambridge, MA: MIT Press.

Lüdtke, J, Friedrich, C. K, De_Filippis, M, & Kaup, B. (2008). Event-related potential correlates of negation in a sentence-picture verification paradigm. *Journal of Cognitive Neuroscience*, *20*(8), 1355–1370. doi:10.1162/jocn.2008.20093

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. doi:10.1037/a0029284

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. doi:10.1016/j.jml.2013.08.001

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. R. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, *33*(7), 1174–1184. doi:10.3758/BF03193221

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye- movements are modulated by reading ability - evidence from Indian low and high literates. *Journal of Eye Movement Research*, *5*(1), 1–10.

Momma, S., Sakai, H., & Phillips, C. (2015). *Give me several hundred more milliseconds: Temporal dynamics of verb prediction*. Talk given at the 28th annual CUNY conference on human sentence processing, Los Angeles, CA.

Nieuwland, M. S, & Kuperberg, G. R. (2008). When the Truth Is Not Too Hard to Handle. *Psychological Science*, *19*(12), 1213–1218. doi:10.1111/j.1467-9280.2008.02226.x

Oishi, H., & Sakamoto, T. (2010). *Immediate interaction between syntactic and semantic outputs: Evidence from event-related potentials in Japanese sentence processing*. Poster presented at the 22nd annual CUNY human sentence processing conference, Davis, CA.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*. doi:10.1016/0028-3932(71)90067-4

Omaki, A., Lau, E., White, I. D., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*, *6*(MAR), 1–30. doi:10.3389/fpsyg.2015.00384

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. doi:10.1016/0749-596X(92)90039-Z

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. A. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, *8*, 89. doi:10.1186/1471-2202-8-89

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

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, *67*, 426–448. doi:10.1016/j.jml.2012.07.003

Ratcliff, R., Hockley, W., & McKoon, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. *Journal of Experimental Psychology. General*, *114*(4), 435–450. doi:10.1037//0096-3445.114.4.435

Smith, N. J., & Levy, R. P. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. doi:10.1016/j.cognition.2013.02.013

Staub, A, & Clifton, C. J. (2006). Syntactic prediction in language comprehension: evidence from either...or. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(2), 425–436. doi:10.1037/0278-7393.32.2.425

Stroud, C., & Phillips, C. (2012). Examining the evidence for an independent semantic analyzer: An ERP study in Spanish. *Brain and Language*, *120*(2), 108–126. doi:10.1016/j.bandl.2011.02.001

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in polish. *Journal of Memory and Language*, *68*(4), 297–314. doi:10.1016/j.jml.2012.12.002

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, *52*, 997–1009. doi:10.1111/psyp.12437

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

Thornhill, D. E., & van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382–392. doi:10.1016/j.ijpsycho.2011.12.007

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Van Berkum, J. J. A. (2009). The neuropragmatics of "simple" utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), *Semantic and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke, UK: Palgrave Macmillan.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(3), 443–467. doi:10.1037/0278-7393.31.3.443

van_Berkum, J. J. A, Hagoort, P, & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657–671. doi:10.1162/089892999563724

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*, 580–591. doi:10.1162/jocn.2008.20054

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

van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, *18*(7), 1181–1197. doi:10.1162/jocn.2006.18.7.1181

van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*(2), 241–255. doi:10.1016/j.cogbrainres.2004.09.002

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

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. doi:10.1162/0898929041920487

Williams, A. (2015). *Arguments in syntax and semantics*. Cambridge: Cambridge University Press.

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. doi:10.1016/j.neuroimage.2012.04.054

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

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975–988. doi:10.1037/a0029206

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, *1*(6), 209–216. doi:10.1007/s13398-014-0173-7.2

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

Ye, Z., & Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research*, *1203*, 103–115. doi:10.1016/j.brainres.2008.01.090

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1–14. doi:10.1037/0278-7393.32.1.1

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes*, *28*(3), 272–302. doi:10.1080/01690965.2011.622905