Electrical brain potentials reveal temporal dynamics of word prediction during language comprehension

Wing-Yee Chow\textsuperscript{1,2,4}, Ellen Lau\textsuperscript{1}, Suiping Wang\textsuperscript{3}, and Colin Phillips\textsuperscript{1}

\textsuperscript{1}University of Maryland College Park

\textsuperscript{2}Basque Center on Cognition, Brain and Language

\textsuperscript{3}South China Normal University

\textsuperscript{4}Corresponding author.

Keywords: Event-related potentials; Language comprehension; Psycholinguistics; Prediction; N400
Abstract

Anticipation of upcoming input has been shown to be critical to the speed and accuracy of many cognitive processes, particularly in domains such as language in which the listener must process an unbroken stream of 3-5 words per second. However, little is known about how the predictions themselves are generated in real time. Here we demonstrate a method for isolating sub-components of predictive processes and examining how they unfold in real time. We took advantage of the word order properties of Mandarin Chinese to isolate the contribution of word order information to prediction by reversing the order of two pre-verbal noun phrase arguments. We manipulated the timing with which the arguments became available relative to the verb in a sentence, and used electroencephalography to inspect the effects of prediction on early processing (350-450ms) in three experiments. We demonstrate that, despite its prominence and unambiguity, word order information informs language prediction only at a measurable delay. These findings call for more attention to the time course with which different sources of information become available to predictive mechanisms in different speaker groups, and open a new window into the subcomponents of predictive processes that are crucial for fast and robust language understanding.
Introduction

Much recent work suggests that generating predictions about the future is a fundamental principle underlying the brain’s operations (Llinás, 2002; Hawkins, 2004; Bar, 2011). After an explosion of work on this topic across the last decade, predictive processes have now been repeatedly demonstrated across domains such as visual (Bar, 2007) and auditory perception (Bendixen, Schröger & Winkler, 2009), motor planning (Wolpert, 1997), and language comprehension (DeLong, Urbach & Kutas, 2005). 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 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.

In the domain of language, previous research has provided much evidence for human comprehenders’ ability to generate linguistic predictions on the fly. For example, when listeners are presented with the beginning of an utterance like ‘The boy will eat…,’ they make more anticipatory eye-movements to a picture of an edible object (e.g., a cake) relative to a neutral utterance such as ‘The boy will move…’ (Altmann & Kamide, 1999). Prediction in language comprehension has been shown to facilitate syntactic (Wicha, Moreno & Kutas, 2004; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005), lexico-semantic (Kutas, Hillyard, 1984; Federmeier & Kutas, 1999), phonological (DeLong et al., 2005), and orthographic processing (Dikker, Rabagliati, Pylkkänen, 2009) of likely upcoming inputs, indicating that comprehenders compute rich expectations about upcoming sounds and words at multiple levels of representation. How the brain achieves this, and what it takes for predictions to be generated in advance of the relevant input, are important open questions.
Event-related brain potentials (ERPs) provide a useful tool for investigating these questions. In particular, the amplitude of the N400 (Kutas & Hillyard, 1980), a negative-going ERP component that starts at around 250ms and peaks at around 400ms post stimulus onset, is modulated by a word’s predictability (Kutas & Hillyard, 1984). 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 linked to semantic memory processing (Kutas & Federmeier, 2000; Lau, Phillips & Poeppel, 2008). The predictability of a word in a given context is commonly operationalized 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). Previous research has repeatedly found that the amplitude of the N400 response during comprehension is inversely related to a word’s cloze probability (Kutas & Hillyard, 1984; DeLong et al., 2005; Federmeier, Wlotko, De Ochoa-Dewald & Kutas, 2007). This relation between the N400 amplitude and a word’s cloze probability has been proposed to reflect facilitated lexico-semantic processing for words that are more expected in context (Federmeier & Kutas, 1999).

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, sentence structure, event schemas, world knowledge, message-level representations) have also been shown to modulate N400 amplitude (Hagoort, Hald, Bastiaansen & Petersson, 2004; Otten, Nieuwland & Van Berkum, 2007; Nieuwland & Kuperberg, 2008; Van Berkum, 2009; Bicknell, Elman, Hare, McRae & Kutas, 2010; Kos, Vosse, van den Brink & Hagoort, 2010; Paczynski & Kuperberg, 2012). 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).
However, very little is known about how the predictions themselves are generated in real time. As prediction can facilitate processing if and only if it precedes the predicted event, it can be seen as a race against time. To take a more vivid analogy, to hit a flying tennis ball we not only need to estimate its trajectory and spin, but we also need to compute those predictions quickly enough in order to move the racquet to the predicted position before the ball gets there. The same is true in the case of language comprehension. Given the complexity of language input and the speed at which it unfolds in time, the mental computations involved in generating linguistic predictions might still be incomplete when relevant input arises. Therefore, as Federmeier and Laszlo (2009) have noted, models of predictive processing must take into account the time course with which different sources of information become available.

Questions about the time course of predictive computations have received rather limited attention in current research. Comprehenders' rapid sensitivity to a word's predictability, as evidenced by their electrical brain responses and eye-movements, have led many to make the simplifying assumption that, at least for young adult native speakers, all contextual information can impact linguistic prediction immediately (Hale, 2001; Levy, 2008; Demberg & Keller, 2008; Smith & Levy, 2013). However, some recent studies have failed to observe predictive processing effects in populations such as older adults and second language learners, and such findings have been proposed to reflect limitations on how quickly contextual information can be processed in different speaker groups (e.g., Federmeier, McLennan, De Ochoa & Kutas, 2002; Federmeier & Kutas, 2005; DeLong et al., 2012; Martin, Thierry, Kuipers, Boutonnet, Foucart & Costa, 2013). To examine potential causes for apparent prediction failures we must identify ways to isolate sub-components of predictive computations and examine how different sources of contextual information impact comprehenders’ prediction as the input unfolds in real time.

In the current study we isolate one specific source of contextual information and examine how it impacts linguistic prediction in real time. In particular, we isolate unambiguously ‘predictive’ structural role information in sentences and demonstrate that its impact on linguistic
predictions critically depends on the timing with which this information becomes available. Information about the structural roles of a verb’s arguments, i.e., which is the subject and which is the object in a clause, is crucial for interpreting thematic relations in a sentence. In many languages this information is unambiguously encoded by word order, e.g., ‘The cop arrested the thief’ vs. ‘The thief arrested the cop’. In a language with a verb-final word order, the contribution of argument role information to language processing can be studied in isolation by reversing the roles of the arguments. For example, in a verb-final 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). Since the words in these two sequences are identical, the difference in verb cloze probability must reflect the contribution of argument role information.

(1) Argument role reversal in a verb-final sentence in Mandarin Chinese:
(a) jīngchá ba xiàotóu zhua-le…
   cop BA thief arrest…
   “The cop arrested the thief.”
(b) xiàotóu ba jīngchá zhua-le…
   thief BA cop arrest…
   “The thief arrested the cop.”

Although previous research has provided much evidence that argument role information is accurately and rapidly used during sentence interpretation, evidence for its impact on linguistic prediction has remained surprisingly scarce. Reversal of the arguments’ roles is readily detected and it modulates the P600 response, a positive-going ERP component that starts at around 500ms and has been associated with error detection and reanalysis (Osterhout & Holcomb, 1992; Hagoort, Brown & Groothuisen; 1993; Coulson, King & Kutas, 1998; Hahne & Friederici, 1999; Kuperberg, 2007; Van de Meerendonk, Kolk, Vissers & Chwilla, 2010). Meanwhile, although argument role reversal can also greatly impact a verb’s cloze probability, many recent studies across different languages (Dutch, Mandarin Chinese, Japanese and English) have reported that comprehenders’ N400 response at the verb is unaffected by role-reversal of the arguments (Kolk,
While some authors have viewed the absence of the N400 as indicating that comprehenders initially form the wrong interpretation of role-reversed sentences (e.g., Kolk et al., 2003; Hoek et al., 2004; Kim & Osterhout, 2005; Kuperberg, 2007), we take these results to show that, despite its clear impact on the verb’s cloze probability, argument role information failed to inform comprehenders’ verb prediction across these cases. Specifically, we hypothesize that argument role information has a delayed impact on comprehenders’ verb prediction.

Here we examined the timing with which argument role information informs linguistic predictions in three ERP experiments using 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 order was manipulated in sentences in which the verb had high vs. low predictability in the canonical condition, and timing was manipulated between experiments. The verb immediately followed the arguments in Experiment 1 (e.g., ‘cop ba thief arrest’); in Experiment 2 a temporal phrase was placed between the arguments and the verb to increase the linear distance between them (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. Following previous work, in all experiments we expect to observe a P600 effect to role-reversals, assuming that comprehenders are sensitive to the fit of the verb to the preceding context. However, the sensitivity of the N400 to role-reversal across these experiments can provide insight into how word order information contributes to predictions about the verb prior to its presentation. If comprehenders’ predictions are immediately impacted by argument order, N400 amplitude should track off-line measures of verb predictability even when the verb immediately follows the arguments. On the other hand, if it takes time for argument order information to impact predictive processes, we expect N400 amplitude to track off-line measures of predictability only when the presentation of the verb is delayed.
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 comprehenders’ offline predictions for the verb (as measured by cloze probability). Since previous studies that examined the effects of argument role reversals did not measure cloze probability, 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 roles on the target verb’s cloze probability. In high-predictability sentences the critical 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%). Both high- and low-predictability 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’ word order while holding other factors such as lexical semantic association constant. If the word order (and thereby the structural roles) of the arguments immediately impacts comprehenders' verb predictions, then argument role reversals should have a bigger effect on the N400 in high-predictability sentences. Alternatively, if the word order of the arguments fails to impact comprehenders' predictions for the verb before it appears in the input, then the N400 should be insensitive to argument role reversals in both high- and low-predictability sentences. In addition, a P600 effect should be observed in both high- and low-predictability sentences if comprehenders detect the implausibility resulted from argument role reversals.

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 (Oldfield, 1971), and had normal or corrected-to-normal vision and no history of neurological disorder. 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-verb-object sentence frames (e.g., ‘cop ba thief …’ and its role-reversed counterpart ‘thief ba cop …’). The 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.

Materials

Across Experiments 1 to 3, all experimental sentences used the highly frequent SOV ba-construction in Mandarin Chinese. This construction requires a transitive verb, and the arguments’ identity and their syntactic roles are evident before the critical verb because the particle ba always follows the subject and immediately precedes the direct object. Role-reversed sentences were created by reversing the order of the pre-verbal arguments in the canonical sentences, both of which were animate. 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. No anomaly was evident before the critical verb. Further, in order to avoid sentence-final wrap-up effects the sentences were extended beyond the critical 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 critical 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. 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.

In addition, across Experiments 1 to 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 these item sets, each target word was paired with two strongly constraining sentence contexts (maximum cloze values: 53% - 100%), such that the target word had a high cloze probability (average = 88%) in one and a low cloze probability (average = 27%) in the other. A critical difference between these sentences and the experimental sentences is that many words from across the prior context contribute to the differential predictability of the target word in these sentences.

Experimental sentences were distributed in 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.

Procedure

Participants were comfortably seated about 100cm in front of a computer screen in a testing room. Sentences were presented one word at a time in a white font (30 pt simplified Chinese characters) on a black background at the center of the screen. Each sentence was preceded by a fixation cross that appeared for 500ms. Each word appeared on the screen for
400ms, followed by 200ms of blank screen. Each word consisted of 1 to 4 characters. The last word of each sentence was marked with a period, followed 1000ms 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 familiarize 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 hours on average.

**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, P4/5, 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 5kΩ. The EEG and EOG recordings were amplified (band-pass filtered at DC - 200 Hz) and digitized online at 1kHz. 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.

**ERP Data Analysis**

Event-related potentials were computed separately for each participant and each condition for the 1000ms after the onset of the critical verb relative to a 100ms baseline preceding the critical verb. ERP data were evaluated for EOG and other artifacts. Trials contaminated by artifacts 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 600-800ms for the P600. We conducted repeated measures ANOVAs on 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). The topographic factors effectively defined nine regions of interest (ROIs): 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, we discuss effects of Predictability only when they interact with the effects of Reversal. Therefore, although all effects involving Predictability or Reversal are reported in the results tables, only effects involving Reversal are discussed in the main text. Further, paired-sample t-tests were conducted to test for the predicted differences in each of the ROIs.

The same procedures were used to examine the effect of cloze probability in the filler sentences for the control comparison. The repeated measures ANOVAs for each time-window fully crossed Predictability (high vs. low cloze probability) with Anteriority and Laterality.

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%.

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. Table 2 shows the results of the statistical analyses.

Argument role reversals elicited a clear P600 effect with the same topographic distribution across both high- and low-predictability sentences. However, there was no clear effect of Reversals on the N400 in either pair, as shown in Figure 1. The omnibus repeated measures ANOVA in the 350-450 ms interval revealed a three-way Reversal × Anteriority × Laterality interaction, but this was not corroborated by the results of the ROI analyses, which revealed no significant differences in any of the regions in either high- or low-predictability conditions. In the 600-800 ms interval, the omnibus repeated measures ANOVA revealed a main effect of reversal, along with Reversal × Anteriority, Reversal × Laterality, and Reversal × Anteriority × Laterality interactions. These observations were corroborated by the results of the ROI analyses, which showed more positive ERPs in the role-reversal condition than in the canonical condition at central and posterior sites in both high- and low-predictability sentences alike. At the same time, no interactions involving predictability and reversal were reliable, suggesting that the size of the P600 effect elicited by role-reversals did not differ between the high- and low-predictability conditions.

**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. The left column of Table 3 shows the results of the statistical analyses.

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 (Federmeier et al., 2007; Van Petten & Luka, 2012). We will return to discuss the relationship between this late frontal positivity and the P600 in the General Discussion.

Discussion

This experiment showed that argument role reversals did not modulate comprehenders’ N400 response to a verb even when they greatly impacted offline predictability of the verb, as quantified by cloze probability. The N400’s complete insensitivity to the difference in predictability elicited by 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 role reversal effects on the N400 in previous work across different languages (e.g., Kolk et al., 2003; Hoeks et al., 2004; van Herten et al., 2005, 2006; Ye & Zhou, 2008; Oishi & Sakamoto, 2010; Chow & Phillips, 2013), although these studies did not quantify the effect of role reversals on off-line predictability. 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 constitutes a clear exception to the classic generalization that the N400 is modulated by a word’s cloze probability. Furthermore, the observation that argument role reversals did not modulate the N400 even when they greatly impacted the verb’s cloze probability (in the high-predictability sentences) suggests that the N400’s insensitivity is not attributable to a weak experimental manipulation.
Also in line with previous findings, we observed that argument role reversals were readily detected in the present study, as evidenced by the participants’ accurate plausibility judgments, and elicited a clear P600 effect. Here we extended previous findings by showing that effects of role reversal on the P600 are obtained regardless of the predictability of the verb. These results show that comprehenders computed an accurate interpretation of the sentences using structural role information and readily detected the implausibility resulted from argument role reversals. This robust sensitivity to argument role information stands in sharp contrast with the insensitivity of the N400 response and indicates that only a subset of language-related processes were insensitive to argument role information.

Further, the cloze probability manipulation in the filler sentences (control comparison) elicited a robust N400 effect. This establishes that the participants in the current experiment displayed standard N400 effects to a common cloze probability manipulation. As this manipulation presents the same target word in completely different sentence contexts across conditions, differences in the target word’s cloze probability cannot be attributed to specific sources of contextual information. Nonetheless, this rules out the possibility that comprehenders simply did not engage in predictive computations in this study, and instead suggests that word order information about the preverbal arguments’ structural roles failed to impact comprehenders’ verb predictions before the verb appeared in the input. The control comparison results also offer an informal way to evaluate an alternative interpretation of the results in the experimental conditions, according to which the verb’s cloze probability failed to modulate the N400 amplitude 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 independently determine the minimum possible size of an N400, it is clear that the average N400 at the targets in the high-predictability sentences was larger than that elicited by the expected words in the filler sentences. Therefore, the lack of an N400 effect in the high-predictability sentences appears unlikely to be due to a floor effect.
What gave rise to the N400’s insensitivity to argument role reversals? One possibility is that comprehenders simply fail to use argument role information when computing predictions about an upcoming verb. In other words, comprehenders may generate predictions for the verb based only on the plausible relations between the preceding nouns, and not based on the argument roles provided by the syntax. Another possibility is that comprehenders did not have sufficient time to incorporate such information into their predictive computations. Notably, one common characteristic among previous role-reversal studies 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 480 ms in Hoeks et al. (2004), 600 ms in Experiment 1 and Chow & Phillips (2013), and 645 ms in Kolk et al. (2003) and van Herten et al. (2005, 2006). Interestingly, a previous study that examined the processing of relative clauses in German has reported a significant N400 effect to argument role reversals when the verb was presented 1100 ms following the onset of the second argument (Mecklinger et al., 1995). We test these competing hypotheses in the following experiment.

Experiment 2

In Experiment 2 we examined the possibility that comprehenders use structural role information to predict upcoming verbs, but that this information has a delayed impact. To this end, we employed the same experimental design as Experiment 1 and increased the linear 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, effectively delaying the presentation of the verb relative to its arguments by 1200 ms. 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 when they clearly impacted the verb’s cloze probability, i.e., in the high-predictability conditions. Alternatively, if comprehenders 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.

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 direct object and the verb (e.g., cop BA thief ZAI last week arrest…), thereby increasing the SOA between the direct object and the verb from 600ms to 1800ms. The choice of temporal 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 to 5 characters long here. 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.
As in Experiment 1, experimental sentences were distributed in two presentation lists. 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 those in 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 artifacts.

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%.

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 high-predictability conditions in Experiment 2. Table 4 shows the results of the statistical analyses.

As in Experiment 1, a clear P600 effect to argument role reversals was observed in both high- and low-predictability conditions. However, in contrast to Experiment 1, argument role reversals in the high-predictability conditions appeared to also elicit an N400 effect. Although the interaction between Reversal and Predictability failed to reach statistical significance in the omnibus repeated measures ANOVAs in the 350-450 ms interval, ROI analyses showed that argument role reversals did impact the ERPs in the high-predictability conditions. Specifically, ROI analyses in the high-predictability conditions revealed that the ERPs were less negative in the canonical condition than in the role-reversed condition in four out of six central and posterior regions. Meanwhile, no reliable difference was observed in any of the regions in the low-predictability conditions. In the 600-800 ms interval the omnibus repeated measures
ANOVA revealed a main effect of Reversal and a Reversal × Anteriority interaction. No reliable interactions involving predictability and reversal were observed for the P600. These observations were corroborated by the results of the ROI analyses, which revealed more positive ERPs in the role-reversed condition than in the canonical condition across all central and posterior sites in both high- and low-predictability conditions.

**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. The middle column of Table 3 shows the result of the statistical analyses.

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 found that, when the presentation of the verb was delayed by 1200ms, an N400 effect of role-reversal emerged in the high-predictability conditions, and not 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 current results suggest that the time course of predictive computations plays a central role in the N400's (in)sensitivity to argument role reversals. Although the interaction between Predictability and Reversal failed to reach statistical significance in the omnibus ANOVA, the results in the ROI analyses clearly showed that the N400 at the target verb was smaller in the canonical condition than in the role-reversed condition in the high-predictability sentences, but did not differ in the low-predictability sentences. Put differently, argument role reversals elicited an N400 effect only when they strongly impacted the target verb’s cloze probability. This is consistent with the hypothesis that the N400’s insensitivity to role-reversals in previous studies shows that comprehenders did not have sufficient time to incorporate information about the arguments’ structural roles into their verb predictions. At the same time, the fact that the effect of role-reversal on cloze probability was reflected in the N400 when additional time was provided argues against the hypothesis that comprehenders simply never use argument role information to inform verb predictions.

These results are also incompatible with proposals that take the N400’s insensitivity to argument role reversal to indicate that, upon encountering the verb, comprehenders assign argument roles according to what is most plausible rather than what the syntax indicates, and thus form a plausible interpretation in both the canonical and role-reversed sentences (e.g., Kolk et al., 2003; Hoek et al., 2004; Kim & Osterhout, 2005; Kuperberg, 2007). Since the arguments and the verb can form a plausible interpretation regardless of the linear distance between them, it remains unclear whether and how these accounts can be reconciled with the findings that the argument role reversals elicited an N400 effect only in the high-predictability sentences in the current experiment.

One unresolved issue, however, concerns the nature of the contribution of the added adverbial phrase in the experimental materials. While the adverbial phrase was added in this experiment to increase the amount of time available for predictive computations, the introduced changes to the experimental materials might have other unintended impacts on predictive
computations. Therefore, in Experiment 3 we conducted a more direct test of this hypothesis by manipulating the distance between the arguments and the verb within a single experiment in which all conditions contained the adverbial phrase.

**Experiment 3**

In this experiment we tested the hypothesis that additional time is required in order for comprehenders to incorporate argument role information into their verb prediction more directly, 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 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. If structural 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 role-reversals when the verb is further away from its arguments, i.e., in the 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. However, in order to reduce the likelihood of a floor effect on the N400 in the short-distance conditions, we avoided using canonical sentences in which the target verb had a very high cloze probability (> 70%).
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 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. The verb was predictable in the canonical sentences (average cloze = 37.4%, range 16.7 - 69.0%) but not in the role-reversed sentences (0% cloze). Only verbs with a cloze probability below 70% in the canonical sentence were used in order to avoid the possibility of a floor effect in the N400. The verb was not always the most likely continuation in the canonical sentences.

Experimental sentences were distributed in 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 randomized 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 Role-reversal. Therefore, although all effects involving Distance and/or Reversal are reported in the results tables, 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 artifacts.

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%.

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. Table 5 shows the results of the statistical analyses.

In Experiment 3, an N400 effect of role-reversal was elicited in the long-distance conditions but not in the short-distance conditions. Omnibus repeated measures ANOVAs in the 350-450 ms interval revealed a Reversal × Distance × Laterality interaction. This observation was corroborated by the results of the ROI analyses, which revealed less negative ERPs in the long-distance canonical sentences than in the long-distance role-reversed sentences across several midline and right central and posterior sites, but no corresponding differences in the short-distance conditions. In the 600-800 ms interval the omnibus ANOVA revealed a main effect of Reversal and a Reversal × Anteriority interaction. No interactions involving distance and reversal were significant for the P600. These observations were corroborated by the results of the ROI
analyses, which revealed more positive ERPs in the role-reversed condition than in the canonical condition across central and posterior sites in both short- and long-distance conditions.

**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 right column of Table 3.

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.

*<Figure 4 and Table 5 about here>*

**Discussion**

Figure 5 presents a summary of the main results of Experiments 1 to 3. The results of Experiment 3 demonstrated that argument roles impacted the amplitude of the N400 when the presentation of the target verb was delayed (in the long-distance conditions) but not when the verb immediately followed its arguments (in the short-distance conditions). Other aspects of the results were also consistent with those of Experiments 1 and 2. First, as in the previous experiments, argument role reversals were readily detected in both short- and long-distance conditions and elicited the same 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 5 about here>*
The N400’s sensitivity to argument role-reversals in the long distance conditions (1800ms SOA) stands in sharp contrast with its insensitivity in the short distance conditions (600ms SOA) 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, and must instead reflect the time at which different pieces of contextual information became available as the sentence unfolded. In particular, the N400’s reemerged sensitivity in the long-distance conditions indicates that access to the critical verb’s meaning was facilitated in the canonical condition when comprehenders had more time to use argument role information in predictive computations. Note that although here we show that argument role information has a delayed impact on prediction, we believe that other kinds of information can impact predictions quite rapidly; we return to this point in the General Discussion.

**General Discussion**

In the current study, we isolated the contribution of argument role information by reversing the order of the preverbal arguments. Then, we manipulated the timing of the arguments relative to the verb by varying the position of a temporal adverbial expression in the sentence. We compared the amplitude of the N400 for the target verb across canonical and role-reversed sentences to examine whether information about the arguments’ structural roles can facilitate semantic access to the target verb via prediction and how its impact develops over time. Results from the three ERP experiments presented here show that the implausibility created by reversal of the arguments in verb-final sentences was readily detected and that it consistently elicited a P600 effect at the verb. However, argument role reversals elicited an N400 effect only when they had a strong impact on the verb’s cloze probability and when there was additional time between the arguments and the verb. These findings license two conclusions.
First, timing is critical to whether word order can impact linguistic predictions. In contrast with its insensitivity in Experiment 1 and in previous studies using role reversals, the N400 became sensitive to cloze probability differences resulting from argument role reversals when the time interval between the arguments and the verb was widened in Experiments 2 and 3. The current data are also consistent with Mecklinger et al.’s (1995) observation that argument role reversals in locally ambiguous German relative clauses elicited a significant N400 effect when the verb was presented 1100 ms following the onset of the second argument. These results indicate that word order information about the arguments’ structural roles cannot immediately impact verb predictions, but that comprehenders’ expectations for the critical verb become stronger in the canonical condition than in the role-reversed condition when they are given more time for prediction. These results provide empirical support for Federmeier & Laszlo’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). Further, the widened time interval (1800ms between the onset of the object and the verb) places an upper limit on the time required for word order information to impact predictive computations. This is to our knowledge the first case in which the time course of a specific component of predictive processing has been isolated, and it opens up a host of questions about how different sources of information are integrated to feed linguistic prediction in real time.

Second, word order successfully impacts comprehension, even in situations where it fails to inform prediction. In the ba-construction in Mandarin Chinese, the structural roles of the preverbal arguments are unambiguously marked by their relative word order. Comprehenders reliably use the word order of the arguments to identify the subject and the object of a sentence and they readily detect the implausibility resulting from argument role reversals, as evidenced by their accurate behavioral judgments and the consistent presence of a P600 effect at the verb. This was seen in all of the current experiments, as well as in previous studies (Kolk et al., 2003; Hoeks et al., 2004; Van Herten et al., 2005, 2006; Chow & Phillips, 2013). Further, this effect
was not modulated by the predictability of the sentences or the distance between the arguments and the verb. In contrast, the same word order information under the same conditions may fail to impact comprehenders' predictions, as seen in Experiment 1. The contrast between comprehenders' overt judgments and P600 sensitivity on the one hand and the N400's insensitivity on the other hand suggests that as long as comprehenders can obtain reliable bottom-up information, their failure to predict upcoming input accurately in time does not preclude successful comprehension, though it may impact the speed and robustness of comprehension.

It is important to note that some have suggested a different explanation for the absence of N400 effects of role reversal observed in previous studies (e.g., Kolk et al., 2003; Hoek et al., 2004; Kim & Osterhout, 2005; Kuperberg, 2007). Under these accounts, N400 amplitude is primarily taken to be an index of semantic/pragmatic plausibility, and the absence of N400 effects of role-reversal is taken to indicate that, upon encountering the verb, comprehenders assign argument roles according to what is most plausible rather than what the syntax indicates, resulting in identical (plausible) interpretations in both the canonical and role-reversed sentences. However, a number of recent results have cast doubts on this explanation (e.g., Brouwer, Fitz, & Hoeks, 2012; Chow & Phillips, 2013). The current results also seem inconsistent with this hypothesis, as it is not clear why extending the time between arguments and verb should change the mechanism by which argument roles are assigned upon encountering the verb.

The role of lexical semantic association

Lexical association is also known to modulate the amplitude of the N400 response, and it is thus important to consider how lexical association between the arguments and the verb might have contributed to the results observed here. In particular, an alternative account of the absence of N400 effects in the short distance condition might be that strong lexical associations between the arguments and verbs resulted in a floor effect in the N400 amplitude such that predictability could not reduce the response further. Under this account, the N400's sensitivity reemerged in
the long-distance conditions not because of a difference in the timing of prediction, but because
the effect of lexical association from the arguments decayed during the adverbial phrase.

Given previous findings that the effect of lexical association decays over time (e.g.,
Ratcliff, Hockley & McKoon, 1985), it seems possible that increasing the distance between
associates and the critical word might make the effects of other factors more easily observable.
However, we consider it unlikely that the N400’s insensitivity to the verb’s cloze probability in
the short-distance conditions can be attributed to a lexical association floor effect. Even though
we cannot independently determine the minimum possible size of an N400, across all three
experiments the average N400 at the targets in the experimental sentences was clearly larger than
that elicited by the expected words in the filler sentences, which suggests that the N400
amplitude for experimental targets was not at floor. As such, we believe that lexical association
alone cannot explain the N400’s complete insensitivity to the verb’s cloze probability when it
appeared immediately following the arguments.

Other cases of “prediction failure”

We demonstrated that 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. 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 predictive computations are not yet complete
when the relevant bottom-up input arises (Hale, 2001; Levy, 2008; Demberg & Keller, 2008;
Smith & Levy, 2013). Further, we argue that careful considerations about timing are crucial for
differentiating between genuine failures to engage predictive mechanisms and cases in which
certain sources of contextual information cannot impact predictions quickly enough to facilitate
bottom-up processing. Such considerations may help us to better understand why predictive
processing effects are absent under certain circumstances (e.g., Chwilla et al., 2011) and in different speaker groups such as older adults (e.g., Federmeier et al., 2002; DeLong et al., 2012), second language learners (e.g., Martin et al., 2013), and patients with schizophrenia (e.g., Ford & Mathalon, 2012).

**Late frontal positivity and the P600**

Although a late positivity was observed in experimental and control comparisons across Experiments 1 to 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.

In fact, a recent meta-analysis by Van Petten and Luka (2012) suggests 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. In the current study, argument role reversals rendered half of the experimental sentences implausible, 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 (e.g., Coulson, King & Kutas, 1998), and/or error corrections in a noisy channel model (e.g., Gibson, Stearns, Bergen, Eddy & Fedorenko, 2013).

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., Federmeier et al., 2007; DeLong et al., 2011; 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 (Kutas, 1993; Federmeier et al., 2007) and/or learning signal from which to update future predictions (Federmeier, Kutas, & Schul, 2010). However, existing results have remained mixed with regard to whether a larger positivity is elicited 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.

**Towards an explicit model of predictive processing**

Here we propose that different sources of information impact linguistic predictions on different time scales. Although information about the arguments’ structural roles may have a delayed impact on comprehenders’ predictions for an upcoming verb, other sources of information about the arguments (e.g., lexical meaning, animacy) may have a more immediate impact (e.g., McRae, Hare, Elman, & Ferretti, 2005). Previous studies have shown that the N400 is modulated by the thematic relatedness between the verb and its argument(s) (Van Herten et al., 2006; Bicknell et al., 2010; Paczynski & Kuperberg, 2012; Chow & Phillips, 2013). For example, a previous study using the same *ba*-construction in Mandarin Chinese has shown that, even when the verb immediately followed its arguments, verbs that were closely related to the arguments elicited a smaller N400 response than those that were less closely related (Chow & Phillips, 2013). Taken together, these findings suggest that lexical semantic information about the arguments can
impact comprehenders’ predictions about an upcoming verb more quickly than information about their structural roles. In subsequent work in English we have found that, depending on the sources of context information being manipulated, cloze probability manipulations of the same magnitude have different effects on the N400 (Chow et al., *in prep*).

Why might information about argument roles have a delayed impact on verb predictions compared to lexical semantic information about the arguments? Just as language inputs unfold in time, different pieces of contextual information and grammatical knowledge must be integrated to feed linguistic predictions. For example, when processing a verb-final sentence, some information about the arguments may feed verb predictions as soon as their lexical representations have been retrieved from long-term semantic memory. By hypothesis, activating the unordered set of arguments *<suspect, police>* could lead to prediction of *arrest* by querying which are the most frequent events to have *suspect* and *police* as participants. On the other hand, it is only after such lexical representations have been integrated into a syntactic structure that the arguments’ structural roles can impact verb predictions. Therefore, it is also possible that predictive computations that involve structural role information are more complex and take longer than those that rely only on lexical information alone. Last but not least, it is likely that the use of different information sources in real time also depends on language-specific properties. For example, structural role information might impact linguistic predictions more quickly when it is encoded in grammatical case (e.g., in German) than when it is encoded by word order, as in English and Chinese.

We propose that some of these questions can be studied systematically with the paradigm we presented in the current study. 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 in future research to examine the time course with which different information sources impact predictive processing and how it might be affected in different populations.

Conclusion

Just as language inputs unfold in time, comprehenders’ access to different sources of contextual information also develops over time. The current study highlights the importance of careful considerations about when different sources of contextual information can impact predictions. This approach can be generalized to study diverse information sources across different populations and will contribute to developing more explicit models of linguistic predictions. By isolating sub-components of predictive processes and how they unfold in time, this approach may prove critical to understanding the success and failures in language comprehension across the lifespan.
References

Tables and Figures

Experimental comparison

Low-predictability

Canonical

Mr. Liu BA parrot

train quite some time.

Role-reversed

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

cop BA thief

arrest (and bring back) to police station.

Role-reversed

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

Herd BA sheep’s body cut-down a lot of wool…

“From the sheep the herdsman collected a lot of wool…”

Unexpected

Boss ZAI last year buy a lot of wool…

“Last year the boss bought a lot of wool…”

Table 1. Experimental conditions and sample materials in Experiment 1 (top) and sample materials for the control comparison in Experiments 1 to 3 (bottom).
Figure 1. Top: Grand average ERPs at centro-parietal electrode CPZ and the topographic distribution of ERP effects in the 350-450 ms and 600-800ms intervals (role-reversed minus canonical) in the low-predictability (left) and high-predictability (right) conditions in Experiment 1. 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 when it appeared immediately following the second argument (600ms SOA). 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 600-800ms 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>
<thead>
<tr>
<th>Factor</th>
<th>df</th>
<th>Omnibus ANOVA</th>
<th>ROI analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>350-450ms</td>
<td>600-800ms</td>
</tr>
<tr>
<td>rev</td>
<td>1,23</td>
<td>&lt; 1</td>
<td>18.78**</td>
</tr>
<tr>
<td>pred</td>
<td>1,23</td>
<td>1.3</td>
<td>1.95</td>
</tr>
<tr>
<td>rev * pred</td>
<td>1,23</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>rev * ant</td>
<td>2,46</td>
<td>1.24</td>
<td>5.97*</td>
</tr>
<tr>
<td>pred * ant</td>
<td>2,46</td>
<td>37.94**</td>
<td>4.23*</td>
</tr>
<tr>
<td>rev * lat</td>
<td>2,46</td>
<td>&lt; 1</td>
<td>4.19*</td>
</tr>
<tr>
<td>pred * lat</td>
<td>2,46</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>rev * pred * ant</td>
<td>2,46</td>
<td>&lt; 1</td>
<td>1.03</td>
</tr>
<tr>
<td>rev * pred * lat</td>
<td>2,46</td>
<td>&lt; 1</td>
<td>1.94</td>
</tr>
<tr>
<td>rev * ant * lat</td>
<td>4,92</td>
<td>3.66*</td>
<td>4.89**</td>
</tr>
<tr>
<td>pred * ant * lat</td>
<td>4,92</td>
<td>3.15*</td>
<td>4.06**</td>
</tr>
<tr>
<td>rev * pred * ant * lat</td>
<td>4,92</td>
<td>&lt; 1</td>
<td>1.2</td>
</tr>
</tbody>
</table>

**ROI analyses**

**Low Predictability**

<table>
<thead>
<tr>
<th>Region</th>
<th>df</th>
<th>F-value</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left anterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Midline anterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>1.29</td>
</tr>
<tr>
<td>Right anterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>1.52</td>
</tr>
<tr>
<td>Left central</td>
<td>23</td>
<td>&lt; 1</td>
<td>1.55</td>
</tr>
<tr>
<td>Midline central</td>
<td>23</td>
<td>1.26</td>
<td>2.08*</td>
</tr>
<tr>
<td>Right central</td>
<td>23</td>
<td>1.18</td>
<td>2.13*</td>
</tr>
<tr>
<td>Left posterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>3.09**</td>
</tr>
<tr>
<td>Midline posterior</td>
<td>23</td>
<td>1.46</td>
<td>3**</td>
</tr>
<tr>
<td>Right posterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>3.73**</td>
</tr>
</tbody>
</table>

**High Predictability**

<table>
<thead>
<tr>
<th>Region</th>
<th>df</th>
<th>F-value</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left anterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Midline anterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>2.73*</td>
</tr>
<tr>
<td>Right anterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>1.2</td>
</tr>
<tr>
<td>Left central</td>
<td>23</td>
<td>&lt; 1</td>
<td>3.59**</td>
</tr>
<tr>
<td>Midline central</td>
<td>23</td>
<td>&lt; 1</td>
<td>3.68**</td>
</tr>
<tr>
<td>Right central</td>
<td>23</td>
<td>&lt; 1</td>
<td>4.62**</td>
</tr>
<tr>
<td>Left posterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>6.07**</td>
</tr>
<tr>
<td>Midline posterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>6.6**</td>
</tr>
<tr>
<td>Right posterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>5.43**</td>
</tr>
</tbody>
</table>

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

Table 2. Omnibus ANOVA F-values and ROI analysis t-values at the target verb for the high- and low-predictability comparisons in Experiment 1.
Experiment 1 | Experiment 2 | Experiment 3
---|---|---
**Omnibus ANOVA**
| df | 350-450ms | 600-800ms | df | 350-450ms | 600-800ms | df | 350-450ms | 600-800ms |
|---|---|---|---|---|---|---|---|---|---|
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** |

**ROI analyses**

- **Left anterior**
  - 23 | 2.27* | 1.43
- **Midline anterior**
  - 23 | 2.43* | 2.06^<1
- **Right anterior**
  - 23 | 2.41* | 2.08*<1
- **Left central**
  - 23 | 3.53** | 2^<1
- **Midline central**
  - 23 | 4.15** | < 1
- **Right central**
  - 23 | 4.61** | 1.74^<1
- **Left posterior**
  - 23 | 4.88** | < 1
- **Midline posterior**
  - 23 | 4.83** | < 1
- **Right posterior**
  - 23 | 5.85** | < 1

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

Table 3. ANOVA F-values and ROI analysis t-values at the critical verb for the control comparisons in Experiments 1-3.
Figure 2. Top: Grand average ERPs at centro-posterior electrode CPZ and topographic distribution of ERP effects in the 350-450 ms and 600-800ms intervals (role-reversed minus canonical) in the low-predictability (left) and high-predictability (right) conditions in Experiment 2. 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. The N400 appeared sensitive to the predictability of the target verb when the linear distance between the verb and its arguments was increased (1800ms SOA). 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 600-800ms 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.
### Omnibus ANOVA

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>350-450ms</th>
<th>600-800ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>rev</td>
<td>1,23</td>
<td>1.37</td>
<td>19.01**</td>
</tr>
<tr>
<td>pred</td>
<td>1,23</td>
<td>9.88**</td>
<td>1.48</td>
</tr>
<tr>
<td>rev * pred</td>
<td>1,23</td>
<td>1.01</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>rev * ant</td>
<td>2,46</td>
<td>&lt; 1</td>
<td>29.63**</td>
</tr>
<tr>
<td>pred * ant</td>
<td>2,46</td>
<td>10.86**</td>
<td>10.22**</td>
</tr>
<tr>
<td>rev * lat</td>
<td>2,46</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>pred * lat</td>
<td>2,46</td>
<td>1.92</td>
<td>1.13</td>
</tr>
<tr>
<td>rev * pred * ant</td>
<td>2,46</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>rev * pred * lat</td>
<td>2,46</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>rev * ant * lat</td>
<td>4,92</td>
<td>1.74</td>
<td>1.47</td>
</tr>
<tr>
<td>pred * ant * lat</td>
<td>4,92</td>
<td>5.05**</td>
<td>5.33**</td>
</tr>
<tr>
<td>rev * pred * ant * lat</td>
<td>4,92</td>
<td>1.35</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

### ROI analyses

#### Low Predictability

<table>
<thead>
<tr>
<th>Area</th>
<th>df</th>
<th>&lt; 1</th>
<th>&lt; 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left anterior</td>
<td>23</td>
<td>1.58</td>
<td></td>
</tr>
<tr>
<td>Midline anterior</td>
<td>23</td>
<td>1.33</td>
<td></td>
</tr>
<tr>
<td>Right anterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Left central</td>
<td>23</td>
<td>&lt; 1</td>
<td>3.45**</td>
</tr>
<tr>
<td>Midline central</td>
<td>23</td>
<td>&lt; 1</td>
<td>3.47**</td>
</tr>
<tr>
<td>Right central</td>
<td>23</td>
<td>&lt; 1</td>
<td>2.96**</td>
</tr>
<tr>
<td>Left posterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>3.97**</td>
</tr>
<tr>
<td>Midline posterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>4.72**</td>
</tr>
<tr>
<td>Right posterior</td>
<td>23</td>
<td>&lt; 1</td>
<td>4.45**</td>
</tr>
</tbody>
</table>

#### High Predictability

<table>
<thead>
<tr>
<th>Area</th>
<th>df</th>
<th>&lt; 1</th>
<th>&lt; 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left anterior</td>
<td>23</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>Midline anterior</td>
<td>23</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>Right anterior</td>
<td>23</td>
<td>1.07</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Left central</td>
<td>23</td>
<td>1.16</td>
<td>2.95**</td>
</tr>
<tr>
<td>Midline central</td>
<td>23</td>
<td>2.41*</td>
<td>2.42*</td>
</tr>
<tr>
<td>Right central</td>
<td>23</td>
<td>2.38*</td>
<td>2.34*</td>
</tr>
<tr>
<td>Left posterior</td>
<td>23</td>
<td>2.68*</td>
<td>4.3**</td>
</tr>
<tr>
<td>Midline posterior</td>
<td>23</td>
<td>2.27*</td>
<td>4.52**</td>
</tr>
<tr>
<td>Right posterior</td>
<td>23</td>
<td>1.58</td>
<td>4.09**</td>
</tr>
</tbody>
</table>

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

Table 4. Omnibus ANOVA F-values and ROI analysis t-values at the target verb for the high- and low-predictability comparisons in Experiment 2.
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; 600ms SOA between the second argument and the verb) or between the direct object and the verb (long distance; 1800ms SOA).
Figure 4. Top: Grand average ERPs at centro-posterior electrode CPZ and topographic distribution of ERP effects in the 350-450ms and 600-800ms 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 the inclusion of an adverbial phrase between the arguments and the verb allows additional time for predictive computations, 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 350-450ms and 600-800ms 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 5. Omnibus ANOVA F-values and ROI analysis t-values at the target verb for the short- and long-distance comparisons in Experiment 3.
Figure 5. The left and center columns show average ERPs at centro-posterior electrode CPZ to the target verb (350-450ms 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 (350-450ms) 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 comprehenders had sufficient time to incorporate information about the arguments’ structural roles into their verb prediction.