Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

A 'bag-of-arguments' mechanism for initial verb predictions

Wing-Yee Chow[1,2,a], Cybelle Smith[1,3], Ellen Lau[1], and Colin Phillips[1]

[1] Department of Linguistics, 1401 Marie Mount Hall, University of Maryland College Park, Maryland, 20742, United States of America

[2] Department of Linguistics, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom

[3] Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel St., Champaign, IL 61820, United States of America

[a] Corresponding author.
Division of Psychology and Language Sciences
University College London
Chandler House
2 Wakefield Street
London
WC1N 1PF
United Kingdom
Telephone: +44 (0) 20 7679 4213
E-mail address: wingyee.chow@ucl.ac.uk (W.Y. Chow).

Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

Abstract

Previous studies have shown that comprehenders use rich contextual information to anticipate upcoming input on the fly, but less is known about how comprehenders integrate different sources of information to generate predictions in real time. The current study examines the time course with which the lexical meaning and structural roles of preverbal arguments impact comprehenders' lexical semantic predictions about an upcoming verb in two event-related potential (ERP) experiments that use the N400 amplitude as a measure of online predictability. Experiment 1 showed that the N400 was sensitive to predictability when the verb's cloze probability was reduced by substituting one of the arguments (e.g., "The superintendent overheard which tenant/realtor the landlord had evicted…"), but not when the verb's cloze probability was reduced by simply swapping the roles of the arguments (e.g., "The restaurant owner forgot which customer/waitress the waitress/customer had served …"). Experiment 2 showed that argument substitution elicited an N400 effect even when the substituted argument appeared elsewhere in the sentence, indicating that verb predictions are specifically driven by the arguments in the same clause as the verb, rather than by a simple 'bag-of-words' mechanism. We propose that verb predictions initially rely on a 'bag-of-arguments' mechanism, which specifically relies on the lexical meaning, but not the structural roles, of the arguments in a clause.

Keywords: Language comprehension; Prediction; Thematic relations; Event-related potentials; N400

Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

Introduction

Just as we are more likely to catch a flying ball if we can anticipate its trajectory, our brain can process information more efficiently if it can anticipate upcoming input ahead of time. In recent years, predictive processes have been argued to play a key role across various cognitive domains such as visual perception, auditory perception, and motor planning (e.g. Bar, 2007; Bendixen, Schröger & Winkler, 2009; Wolpert, 1997). In the domain of language, much previous research has argued that comprehenders can generate rich predictions about likely upcoming inputs on the fly (e.g. Altmann & Kamide, 1999; Federmeier & Kutas, 1999; Wicha, Moreno & Kutas, 2004; DeLong, Urbach & Kutas, 2005; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005; Dikker, Rabagliati, Pylkkänen, 2009). For example, Altmann and Kamide (1999) found that listeners make more anticipatory eye-movements to a picture of an edible object (e.g., a cake) when they are presented with the beginning of an utterance like "The boy will eat …" relative to a neutral utterance such as "The boy will move …." In a later study, Van Berkum and colleagues (2005) measured participants' event-related brain potentials (ERPs) as they listened to Dutch sentences such as "The burglar had no trouble locating the secret family safe. Of course, it was situated behind a big_{NEU} / big_{COM} but unobtrusive $painting_{NEU}$ / $bookcase_{COM}$." and found that, even prior to the onset of the noun, the adjective (e.g., big_{COM}) elicited an early positivity when its grammatical gender was inconsistent with that of the predicted noun (e.g., $painting_{NEU}$), suggesting that comprehenders pre-activated the most likely noun continuation and its gender.

These findings suggest that comprehenders can rapidly integrate various sources of contextual information to compute predictions on the fly, and 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, as we detail below, recent evidence suggests that some contextual information is incorporated into

predictions more quickly than others. Building explicit and accurate models of linguistic prediction will thus require us to isolate different sources of contextual information and examine how each of them impacts predictive computations in real time. In the current study we examine how comprehenders use the meaning and structural position of the preceding words, in conjunction with their event-based knowledge, to compute lexical semantic predictions about an upcoming verb in a sentence. We find evidence that the meaning of the arguments of the upcoming verb (rather than all preceding nouns) is rapidly used to predict that verb, but that the structural position of those arguments is used in bottom-up integration but is not as quickly incorporated into predictions.

Here we will take for granted that comprehenders' predictions can be modulated by different sources of information in the input. For example, the verb "evict" is improbable in both "Which cat did the landlord evict?" and "Which landlord did the tenant evict?" but the improbability calls on different types of information in the two cases. In the first case, an evicting event is improbable given the event participants (a landlord and a cat). In the latter case, even though an evicting event may occur between a landlord and a tenant, it is very improbable given their roles in the sentence. However, as language processing is highly incremental in nature, different sources of predictive information may become available and impact predictive computations at different points in time. Further, as our goal is to study the mechanisms through which comprehenders pre-activate stored representations (e.g., of words or features) in long-term semantic memory in anticipation of upcoming inputs, we use 'prediction' to refer to pre-activation of stored representations before the bottom-up input arises and we make no presupposition regarding the nature of predictive mechanisms (e.g., whether they are autonomous or controlled).

Previous literature

Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

Event-related brain potentials (ERPs) provide a useful tool for studying the cognitive processes that underlie real-time language comprehension. Of particular interest here is 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). 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). Crucially, previous research has repeatedly found that the amplitude of the N400 response to a word during comprehension is inversely related to that word's offline predictability (Kutas & Hillyard, 1984; DeLong et al., 2005; Federmeier, Wlotko, De Ochoa-Dewald & Kutas, 2007; Wlotko & Federmeier, 2012). This has been proposed to reflect facilitated long-term semantic memory access due to pre-activation of likely upcoming words or semantic features (Federmeier & Kutas, 1999). The predictability of a word in a given context is commonly operationalized as its cloze probability, which is the proportion of trials on which speakers continue the sentence context with that word in an untimed sentence fragment completion task (Taylor, 1953). Many factors that can affect a word's cloze probability (e.g., negation, sentence structure, event schemas, world knowledge, message-level representations) also modulate the size of the N400 (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).

However, recent work suggests that not all contextual information impacts comprehenders' predictions immediately. For example, in a visual-world eye-tracking study, Kukona and colleagues (Kukona, Fang, Aicher, Chen & Magnuson, 2011) reported eye-movement evidence which suggests that listeners cannot immediately use information about the thematic role of a preceding argument to predict an upcoming argument. The authors examined listeners' eye fixations on displays containing

verb-related agents and patients (e.g., policeman and crook) as they listened to active sentences such as "Toby arrests the crook" and passive sentences such as "Toby was arrested by the policeman." Upon hearing the verb, if listeners can rapidly use the thematic role of the first argument to predict the second argument, then they should be more likely to anticipatorily fixate on plausible patients in active sentences and on plausible agents in passive sentences. However, the predicted pattern was observed in passive sentences only; listeners were nearly as likely to fixate agents as patients upon hearing the verb in active sentences. As passive sentences provide additional syntactic cues and time between verb and target noun compared to active sentences, these results may be taken to suggest that thematic role information has a delayed impact on comprehenders' predictions.

Relatedly, recent ERP findings suggest that information about the structural roles of preverbal arguments may also fail to immediately impact comprehenders' predictions for an upcoming verb. In a verb-final clause, reversing the arguments (e.g., cop_{SUBJ} $thief_{OBJ}$ vs. $thief_{SUBJ}$ cop_{OBJ}) can greatly impact the cloze probability of the verb (e.g., arrest). However, many studies across different languages have found that argument role reversals do not modulate the N400 at the verb at all (Kolk, Chwilla, Van Herten & Oor, 2003; Hoeks, Stowe & Doedens, 2004; Van Herten, Kolk & Chwilla, 2005; Van Herten, Chwilla & Kolk, 2006; Chow & Phillips, 2013). These results constitute clear exceptions to the well-established generalization that the N400 is modulated by a word's cloze probability.

At the same time, it is important to note that while information about the arguments' roles appears not to impact the ease of accessing the verb in long-term semantic memory (as indexed by the N400), other evidence indicates that argument role information is correctly used in interpreting thematic relations at the verb. The implausibility resulting from argument role reversals is readily detected (e.g., Chow & Phillips, 2013), and previous studies have routinely shown that it elicits a larger late positivity at the verb (a P600 effect; Osterhout & Cohen, 1992). Even though the functional significance of the P600 effect is still debated (e.g., Van Herten et al., 2006; Kuperberg, 2007;

Bornkessel-Schlesewsky & Schlesewsky, 2008; Kim & Sikos, 2011; Brouwer, Fitz & Hoeks, 2012; Gibson, Stearns, Bergen, Eddy & Fedorenko, 2013), the fact that argument role reversals consistently elicited a P600 effect shows that comprehenders incrementally process information about the arguments' roles to interpret these sentences.

These findings are consistent with the hypothesis that it takes time for argument role information to impact comprehenders' lexical semantic predictions about an upcoming verb (Chow & Phillips, 2013). We directly tested this hypothesis in a recent ERP study using the Subject-Object-Verb (SOV) *ba*-construction in Mandarin Chinese (Chow, Lau, Wang & Phillips, submitted). We manipulated the structural roles of the preverbal arguments by reversing their order. We further manipulated the linear distance between the arguments and the verb by placing a temporal adverbial expression (e.g., yesterday afternoon) either at the beginning of the sentence or between the arguments and the verb. This effectively increased the time elapsed between the onset of the second argument and the verb from 600 ms to 1800 ms. An example is shown in (1).

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

Crucially, we found that the N400 at the verb became sensitive to the cloze probability difference resulting from argument role reversals only when the verb was further downstream from its arguments. Further, the effect of timing was found only when the verb was predictable. In cases where the role reversal is clearly implausible but the verb was not predictable (e.g., (2)), adding time did not make a difference (Chow et al., submitted).

(2) *Lao-liu ba yingwu zai qunian-xiatian xunlian-le* …
Mr. Liu BA parrot ZAI last summer train
"Mr. Liu trained the parrot last summer."

These results were corroborated by a recent study in Japanese by our colleagues (Momma, Sakai & Phillips, in prep.), who found that merely increasing the SOA between an argument and a verb in simple two-word Subject-Verb and Object-Verb sentences revived the N400's sensitivity to argument role reversals. Taken together, these results suggest that, contrary to the simplifying assumption that all contextual information can impact linguistic prediction immediately, the structural roles of preverbal arguments can impact comprehenders' lexical semantic prediction about an upcoming verb only at a measurable delay.

On the other hand, other ERP work suggests that not all aspects of verb prediction are slow. Cloze probability differences that result from manipulations of other contextual information have been reported to elicit an N400 effect in verb-final languages such as Dutch (e.g., Hoeks et al., 2004; Van Herten et al., 2006) and Japanese (Oishi & Sakamoto, 2010), as well as in non-verb-final languages such as English (Garnsey, Tanenhaus & Chapman, 1989) and Chinese (Chow & Phillips, 2013). For example, in a classic study by Garnsey and colleagues (1989), comprehenders read sentences with an Object-Subject-Verb (OSV) embedded object question like "The businessman knew which customer/article the secretary called…," and they showed a smaller N400 at the embedded verb (e.g., called) when the fronted object was animate (e.g., customer) than when it was inanimate (e.g., article). These findings suggest that comprehenders can use some information about the arguments for verb predictions rather quickly.

Taken together, these results suggest that comprehenders can use both the meaning and structural roles of preceding words to pre-activate likely upcoming verbs (or their lexical semantic features), but the predictive processes involved are initially sensitive to only a subset of this information. In this study we investigate exactly which information is recruited in the initial stage of

verb prediction and why. One possibility is that all aspects of structural verb-argument computation have a relatively slow impact on verb predictions, and cases in which the verb shows immediate N400 sensitivity simply reflect rapid lexical priming from all the preceding words in the sentence. We will call this the 'Bag-of-words' hypothesis. Another possibility is that comprehenders can identify the arguments in a clause before their structural roles and that they initially rely on their lexical meaning (but not their structural roles) to compute lexical semantic predictions about a likely upcoming verb. We will call this the 'Bag-of-arguments' hypothesis.[1] Figure 1 presents a schematic representation to highlight the difference between the Bag-of-arguments hypothesis and the Bag-of-words hypothesis.

<Figure 1 about here>

Despite apparent similarities, these hypotheses posit distinct processing mechanisms and have rather different implications for theories or models of linguistic prediction. In particular, it is only under the Bag-of-arguments hypothesis that initial verb predictions are taken to be informed by the clausal structure of the sentence. In order to identify the arguments of the embedded verb, this predictive mechanism must rely on a syntactic representation of the sentence that (at least) marks clause boundaries. In other words, even though the Bag-of-arguments hypothesis posits that verb prediction is initially not sensitive to the structural roles of the arguments, it requires at least a rudimetary chunking of the sentence. On the other hand, under the Bag-of-words hypothesis, the lexical meaning of all preceding words are expected to impact comprehenders' predictions regardless of their structural positions. Therefore, distinguishing between these competing hypotheses will help inform theories about the role of syntactic structure in real-time lexical semantic prediction.

[1] We acknowledge that there is a distinction between being an argument of a verb vs. being in the same clause as the verb. However, the present study was not designed to address this distinction. We will return to this in the General Discussion.

The present study

We conducted two ERP experiments to examine how the lexical meaning and structural roles of preceding words impact comprehenders' on-line lexical semantic predictions about an upcoming verb in verb-final clauses. We used embedded object questions in English, which has an OSV word order and presents both the lexical meaning and structural roles of arguments prior to the verb. We manipulated argument role information by reversing the order of the preverbal arguments (argument role reversal). In addition, we changed the lexical semantic information supplied by the arguments by substituting one of the preverbal arguments with another discourse-compatible noun (argument substitution). Using the amplitude of the N400 response to a word as an index of the ease of accessing that word in long-term semantic memory (e.g., Deacon et al., 2000; Kutas & Federmeier, 2000; Lau, Phillips & Poeppel, 2008), in Experiment 1 we examined the extent to which comprehenders' initial predictions about an upcoming verb (or its lexical semantic features) are sensitive to the lexical meaning and structural roles of the arguments by comparing the effects of argument role reversals and argument substitution on N400 amplitudes at the verb. In Experiment 2 we contrasted the Bag-of-arguments and Bag-of-words hypotheses by examining the effect of argument substitution in sentences that contained an identical vs. a different set of words.

Experiment 1

In this experiment we examined whether initial verb predictions are equally sensitive to the lexical meaning and structural roles of the arguments. If information about the arguments' lexical meaning has a more immediate impact on comprehenders' verb predictions than their structural roles, then there should be a time interval following the presentation of the arguments during which comprehenders' verb predictions are sensitive to the arguments' lexical meaning but not to their structural roles.

We isolated the contribution of these two sources of contextual information on verb predictions by manipulating the sentence context in two ways, both of which resulted in a change in the target verb's cloze probability, resulting in a fully crossed 2 (argument role reversal vs. argument substitution) by 2 (high vs. low cloze probability) experimental design (Table 1). The argument role reversal manipulation isolated the contribution of argument role information by reversing the order of the arguments of the target verb while holding everything else constant. The argument substitution manipulation isolated the contribution of the lexical meaning of the arguments by substituting one of the arguments (the fronted object) with a different but discourse-compatible noun, while holding the other argument (the embedded subject) and the target verb constant. A noun was considered to be discourse-compatible in a given context if it is semantically and pragmatically congruous up to the point when the target verb appears.

Crucially, the argument role reversal and substitution manipulations are matched in a number of ways in order to provide a fair comparison between the contribution of the arguments' structural roles and their lexical meaning to comprehenders' on-line verb predictions. First of all, we ensured that argument role reversals and substitution had the same impact on the verb's offline predictability by carefully matching the cloze probability differences between high and low cloze conditions across both types of sentences. In addition, we held constant the distance between the arguments and the target verb across conditions by using the same sentence structure (embedded object question) in all experimental items. Further, we avoided potential animacy confounds by using only animate preverbal arguments in all experimental sentences.

<Table 1 about here>

Based on previous findings, we expect the N400 response at the target verb to be insensitive to cloze probability differences that resulted from argument role reversals. However, if

comprehenders can quickly use the lexical meaning of the arguments for verb predictions, then we should observe an N400 effect at the target verb in the argument substitution conditions, despite the fact that the size of the target cloze probability difference in the argument substitution conditions was highly comparable to that in the argument role reversal conditions. In addition, based on Van Petten and Luka's (2012) observation that semantically incongruous words elicited a P600 effect in addition to an N400 effect in one-third of the 64 published comparisons surveyed, we expect that the target verbs may elicit a P600 effect as they were less semantically plausible in the low cloze conditions than in the high cloze conditions.

Methods

Participants

Twenty-four students (7 female, mean age = 21.9 years, range 18 - 29 years) from the University of Maryland, College Park participated in the current study. All participants were native speakers of English, 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 10 USD/hour for their participation.

Materials

Following Garnsey et al. (1989), the experimental materials made use of embedded object questions, as in "The restaurant owner forgot which customer the waitress had served during dinner yesterday." Unlike the canonical Subject-Verb-Object (SVO) word order in English, this construction has an Object-Subject-Verb (OSV) word order. Therefore, the arguments' identity and their syntactic roles are evident before the verb. Further, the target verb was always preceded by the auxiliary "had." The auxiliary not only served as a pre-target baseline that was matched across conditions, but also

disambiguated the structural assignment of the target word by ruling out alternative interpretations that would have been available otherwise (e.g., "The restaurant owner forgot which waitress the customer (who was) served tipped generously."). Upon the presentation of the embedded subject *the waitress,* although the fronted phrase *which customer* could in principle turn out to be an indirect object, or the subject of a further clause, much evidence from filler-gap dependency processing indicates that readers initially analyze such phrases as direct objects (e.g., Garnsey et al., 1989; Hickok et al., 1992; Boland et al., 1995).

A sample set of experimental items is shown in Table 1. The experimental materials were developed in three stages. First we created a total of 220 pairs of sentence frames up to the auxiliary 'had'. Half of the sentence frames had the same embedded arguments but in reversed structural roles (argument role reversal), while the other half had the same embedded subject but different fronted objects (argument substitution). To determine the cloze probability of different verbs in these sentence frames, we conducted a norming procedure with native English speakers using the Amazon Mechanical Turk (AMT) online marketplace. Sentence frames were divided into four lists of 110 items each. Each list was completed by 30 participants. In accordance with standard cloze norming procedures, participants were asked to read each sentence frame and to supply a word or phrase that they expected to see next. In almost all trials participants responded with a verb.

From the resulting database, 60 pairs of argument role reversal sentences and 60 pairs of argument substitution sentences were selected for use in the ERP experiment. For each pair we selected a target verb that had a cloze value of at least 13% in one version (high cloze condition) and 0% in the other version (low cloze condition). In an argument role reversal item, the two sentences had an identical verb-argument triplet and differed only in the order of the arguments. Meanwhile, in an argument substitution item, the two sentences had a different fronted object but the embedded subject and target verb were identical. All experimental items were active sentences and a majority of

the target verbs had <agent, patient> arguments. A complete set of the materials is available in the Supplementary Materials.

The manipulation of context (argument role reversal vs. substitution) was fully crossed with target cloze probability (high vs. low). In the high cloze conditions the target verbs had an average cloze probability of 25% (range = 13% - 53%) in the argument role reversal sentences and 28% (range = 13% - 77%) in the argument substitution sentences. Further, the target verbs' cloze probability in the high cloze conditions was matched by the first three quartiles across the argument role reversal and substitution sentences (first quartile: 19% and 17%; second quartile: 23% and 23%; third quartile: 30% and 33%). As the target verbs always had zero cloze probability in the low cloze conditions, the differences between the target cloze probability in the high vs. low cloze conditions were closely matched between argument role reversal and argument substitution sentences. While low cloze sentences tended to be less semantically plausible than high cloze sentences, they varied along a continuum of plausibility and some might be considered more plausible than others. 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.

While argument role reversal affected only the structural roles of the arguments, argument substitution resulted in sentences that differed by one word (the fronted object; e.g., "tenant" vs. "realtor") between the high vs. low cloze conditions. We employed Latent Semantic Analysis (LSA; Deerwester, Dumais, Furnas, Landauer & Harshman, 1990; Landauer & Dumais 1997) to assess the extent to which the target verb was lexically associated with the fronted object in the high vs. low cloze argument substitution conditions. LSA is a technique akin to factor analysis, in which words are placed in a high-dimensional space on the basis of word-document co-occurrence frequencies.[2] We

[2] The space is constructed by creating a table of the log-entropy normed frequencies with which stemmed words occur in a collection of documents, performing singular value decomposition (SVD) on this table, and

obtained Semantic Similarity Values (SSVs; max = 1 for identical word pairs) between the fronted object and the target verb for each experimental sentence from a readily available LSA model based on the TASA corpus (available at lsa.colorado.edu), which contains general reading material that a student would be expected to have read up through their first year of college. Term-to-term pairwise comparisons on 300 factors revealed that the verb was mildly associated with the fronted object in both high and low cloze conditions, with mean SSVs of .22 (SD =.2) and .12 (SD = .12) in the high and low cloze conditions respectively.

Sixty pairs of argument role reversal sentences and 60 pairs of argument substitution sentences were distributed in two presentation lists, such that only one member of each pair appeared in each list (30 per condition). In addition, each list also contained 140 filler sentences. Sixty of these filler sentences were included to test the effects of cloze probability in strongly constraining sentences (cloze probability of high cloze words > 75%) and served as a control comparison for the current experiment. These sentences were adapted from a previous experiment examining the effects of cloze probability: each sentence context had a high cloze target and a low cloze target as a continuation, and the high cloze target in one context appeared as the low cloze target in another context. These sentences were distributed in two presentation lists such that each participant read 30 sentences with a high cloze target and 30 sentences with a low cloze target. An additional 80 unrelated filler sentences with similar length and structural complexity were included in both presentation lists. The overall high-to-low cloze ratio in each presentation list was 1:1.

Procedure

reducing the dimensionality of the SVD matrices. For a more detailed introduction to LSA, including a discussion of its limitations as a measure of semantic similarity, please see Landauer, Foltz and Laham (1998).

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 on a black background at the center of the screen. Each sentence was preceded by a fixation cross that appeared for 500 ms. Each word appeared on the screen for 300 ms, followed by 230 ms of blank screen (i.e., 530 ms SOA). The last word of each sentence was marked with a period, followed 1000 ms later by the question "Is this sentence plausible?" 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. Participants were instructed to respond based on whether the sentence depicts "something that would happen normally." 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 five blocks of 52 sentences each, with short pauses in between. The order of the blocks and the sentences within each block were randomized across participants. Including set-up time, an experimental session lasted around two hours on average.

EEG Recording

EEG was recorded continuously from 29 AgCl electrodes mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, and O1/2. Scalp electrodes were referenced online to the left mastoid and re-referenced to the average of both mastoids offline. The electro-oculogram (EOG) was recorded at four bipolar 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 and digitized online at 1kHz with a bandpass filter of 0.1-100 Hz.

Behavioral Data Analysis

Plausibility judgment data in the experimental conditions were analyzed using mixed-effects logistic regression (Jaeger, 2008) with context, cloze and their interaction as fixed effects and by-item and by-subject random intercepts.

ERP Data Analysis

All trials were evaluated individually for EOG or other artifacts. Trials contaminated by artifacts were excluded from the averaging procedure. This affected 9.7% of experimental trials. Event-related potentials were computed separately for each participant and each condition for the 1000 ms after the onset of the target word relative to a 100 ms pre-stimulus baseline. Statistical analyses on average voltage amplitudes were conducted separately for two time windows: 300-500 ms for the N400 and 700-900 ms for the P600 / late positivity. We conducted repeated measures ANOVAs on time-window average ERPs at 18 electrodes, fully crossing context (argument role reversal vs. substitution) and cloze probability (high vs. low) with anteriority (anterior vs. central vs. posterior) and laterality (left vs. midline vs. right). Data from the 60 filler sentences used for the control comparison were analyzed separately, fully crossing cloze probability with anteriority and laterality. 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 context was manipulated between different item sets in the current experiment, we discuss effects of context only when they interact with the effects of cloze probability. Therefore, although all effects involving context or cloze probability are reported in the results table, only effects involving cloze probability are discussed in

the main text. Significant interactions between cloze probability and context were resolved by comparisons within each level of context.

Further, using the *bootES* package (Gerlanc & Kirby, 2013) we computed standardized effect sizes (Pearson's *r*) and their bootstrap confidence intervals (95% CI) for the contrasts between the high and low cloze conditions in both the experimental and filler items using the mean amplitude at 4 midline central-posterior elecrodes (CZ, CPZ, PZ, OZ) in the N400 and P600 time intervals. An advantage of expressing the effect size of a contrast using Pearson's *r* is that the effect is scaled to have an absolute magnitude between 0 and 1, with the sign indicating the direction of the differences.

Results

Plausibility Judgments

As shown in Table 2, participants judged sentences in the high cloze conditions to be plausible at a much higher rate than those in the low cloze conditions. Mixed-effects logistic regression revealed main effects of both factors (context: β = -0.22, p(Wald) < .0001; cloze: β = -1.60, p(Wald) < .0001) and no interaction between them: sentences in the argument substitution conditions were judged 'plausible' around 6% more often than in the argument role reversal conditions; sentences in the high cloze conditions were judged 'plausible' around 60% more often than in the low cloze conditions; the effect of cloze probability did not differ between role reversal and substitution sentences. The non-zero values in the plausibility judgments in the low cloze conditions reflected the fact that low cloze sentences varied along a continuum of plausibility.

<Table 2 about here>

Event-related Potentials (experimental comparisons)

<Figure 2 about here>

<Figure 3 about here>

Figure 2 shows the grand average ERPs at the target word and the topographic distribution of the effects in the 300-500 ms and 700-900 ms intervals in both experimental and control comparisons. Figure 3 shows the grand average ERPs at the target word across all 9 ROIs in the argument role reversal and substitution conditions. Visual inspection of the data indicates that an N400 effect was elicited only in the argument substitution sentences, in which the N400 was smaller for high cloze than low cloze target verbs. Meanwhile, low cloze target verbs elicited a larger late positivity than high cloze target verbs (a P600 effect) in both pairs of conditions.

These observations were confirmed by the statistical analyses. Results from the omnibus ANOVAs are presented in Table 3. The standardized effect size (Pearson's *r*) and 95% confidence interval for the contrasts between the high and low cloze conditions at 4 midline central-posterior electrodes in the N400 and P600 time intervals are presented in Table 4.

In the 300-500 ms interval, the omnibus ANOVA revealed a significant context × cloze probability interaction, along with a main effect of cloze probability and a cloze probability × laterality interaction. Paired-sample t-tests within each level of context on ERPs averaged across all 18 electrodes revealed a significant effect of cloze probability in the Argument Substitution conditions ($t(23) = 2.87$; $p < .01$), but not in the Argument Role Reversal conditions ($t < 1$). The N400 was reduced for high cloze targets compared with low cloze targets in the Argument Substitution conditions, but it was not modulated by cloze in the Argument Role Reversal conditions. In the 700-900 ms interval the omnibus ANOVA revealed a significant main effect of cloze probability. Low cloze targets elicited a larger late positivity than high cloze targets in both Argument Substitution and Role Reversal conditions.

<Table 3 about here>

<Table 4 about here>

Event-related Potentials (control comparison)

The grand average ERPs at the high and low cloze target words for the control comparison are shown in the lower part of Figure 4. The control comparison showed that comprehenders display clear N400 sensitivity to the standard cloze manipulation in the filler sentences. The N400 was reduced for high cloze compared to low cloze words, and low cloze words elicited a larger late positivity than high cloze words.

These observations were confirmed by the statistical analyses. Results from the omnibus ANOVAs are presented in Table 5. The standardized effect size (Pearson's *r*) and 95% CI for the contrasts between the high and low cloze conditions at 4 midline central-posterior electrodes in the N400 and P600 time intervals are presented in Table 6. In the 300-500 ms interval the omnibus ANOVA revealed a main effect of cloze probability, two-way cloze probability × anteriority and cloze probability × laterality interactions, and a three-way cloze probability × anteriority × laterality interaction, due to the fact that the effect was largest at the midline central and posterior sites. In the 700-900 ms interval the omnibus ANOVA revealed a main effect of cloze probability, due to the fact that low cloze words elicited a larger late positivity than high cloze words across the scalp.

<Table 5 about here>

<Table 6 about here>

Discussion

This experiment was designed to determine whether comprehenders' initial lexical semantic predictions about an upcoming verb are equally sensitive to the lexical meaning and structural roles of

the arguments. We reasoned that if a type of information can impact initial verb predictions, then it should impact the ease with which the verb can be accessed from long-term semantic memory. We operationalized ease of long-term semantic memory access as the extent to which the amplitude of the N400 response is reduced (less negative, indicating facilitated access). We found that, even though argument substitution and argument role reversal impacted the target verb's offline predictability (cloze probability) to roughly the same extent, only argument substitution impacted the N400 at the verb during comprehension. We take this to show that access to long-term semantic memory, as indexed by the N400, was facilitated for the high cloze target verbs compared to the low cloze target verbs in the argument substitution conditions but not in the argument role reversal conditions. These results suggest that, by the time the target verb was presented, the lexical meaning of the arguments, but not their structural roles, have impacted comprehenders' lexical semantic predictions about the verb.

However, since the argument substitution manipulation in the current experiment resulted in sentence contexts that differed by one word (the fronted object), these results are ambiguous as to whether initial verb predictions distinguish arguments from non-arguments in the sentence context. As illustrated in Figure 1, the Bag-of-words hypothesis posits that the meaning of all preceding words (arguments and non-arguments alike) contributes to comprehenders' predictions equally. It is only under the Bag-of-arguments hypothesis that arguments are distinguished from other words in the context and impact comprehenders' initial verb predictions selectively. Unlike a Bag-of-words mechanism, which can operate even in an unstructured list of words, a Bag-of-argument mechanism requires at least a rudimentary chunking of the sentence (e.g., identifying clause boundaries). Therefore, in order to better understand the predictive mechanisms at work during online language comprehension, it is important to distinguish these hypotheses and examine the extent to which initial verb predictions are sensitive to the syntactic properties of the input.

In the meantime we found evidence that participants reliably detected the implausibility that resulted from both argument role reversal and substitution. Sentences with a high cloze target verb were judged 'plausible' over 85% of the time, while those with a low (zero) cloze target verb were deemed 'plausible' much less often (< 30%). Despite the fact that the low cloze verb can form a plausible thematic relation with its arguments in the role reversal condition but not in the substitution condition, there were actually slightly more 'implausible' judgments to low cloze sentences in the role reversal condition (76%) than in the substitution condition (69%). Further, low cloze target verbs elicited a larger late positivity (P600) than high cloze target verbs in both cases. This is in line with previous findings (e.g., Chow et al., submitted) and shows that neither the comprehenders' plausibility judgments nor their P600 response was affected by the presence of a plausible thematic relation among the verb and its arguments (regardless of their roles). Therefore, while structural role information was not immediately incorporated into verb predictions, the judgment data and P600 responses indicate that this information was immediately used to interpret thematic relations, and are generally in line with the view that language comprehension is incremental.

Taken together, these results suggest that while participants used both the lexical meaning and structural roles of the arguments to compute an accurate interpretation of the sentences, their initial verb predictions are sensitive to the lexical meaning, but not the structural roles, of the arguments. These results, however, cannot rule out the possibility that initial verb predictions do not distinguish arguments from other words in the context and simply rely on the meaning of all the words in the preceding context (the Bag-of-words hypothesis). We test these competing hypotheses in the following experiment.

Experiment 2

In this experiment we contrasted the Bag-of-arguments and Bag-of-words hypotheses by studying the effect of argument substitution in sentences that contained an identical vs. a different set of words. To this end, we realized argument substitution in two different ways, both of which resulted in a change in the target verb's cloze probability (Table 7). The "different words" argument substitution manipulation is identical to the argument substitution manipulation in Experiment 1, where the fronted object was substituted with a different but discourse-compatible noun. This served as an attempt to replicate the findings in the argument substitution conditions in Experiment 1. In addition, we introduced a "same words" argument substitution manipulation, which substituted the fronted object without changing the lexical content of the sentence. This was achieved by exchanging the fronted object and the main clause subject to create the low cloze condition. As illustrated in the bottom half of Table 7, "neighbor" was the fronted objected in the high cloze condition and the main clause subject in the low cloze condition, while the opposite was true for "exterminator". This resulted in a 2 × 2 within-participants design, fully crossing context (different-words vs. same-words argument substitution) and target cloze probability (high vs. low).

<Table 7 about here>

As in the argument substitution conditions in Experiment 1, both hypotheses predict an N400 effect in the different-words conditions. However, they make different predictions about the effect of argument substitution on the N400 in the same-words conditions. If comprehenders do not immediately distinguish the arguments from other words in the context and instead generate predictions about the upcoming verb in a uniform, 'bag-of-words' fashion (the Bag-of-words hypothesis), then the ease of accessing the target verb in long-term semantic memory should not differ between the high and low cloze sentences as long as the verb is preceded by an identical set of words, as in the same-words conditions in the current experiment. Therefore, under the Bag-of-words

hypothesis, we expect to see an interaction between context and cloze probability in the N400 time interval: an N400 effect is expected in the different-words conditions but not in the same-words conditions.

Alternatively, if comprehenders quickly identify the arguments of the current clause and use their lexical meaning to predict the upcoming verb (the Bag-of-arguments hypothesis), then the ease of accessing the verb in long-term semantic memory should crucially depend on what the arguments are. In the current experiment, the high and low cloze sentences contained different sets of arguments in both the same-words and different-words conditions. For example, in the example in Table 7, the arguments are "neighbor" and "landlord" in the high cloze sentences, but "exterminator" and "landlord" in the low cloze sentences. Therefore, the ease of accessing the verb in long-term semantic memory should be impacted in both cases. In other words, under the Bag-of-arguments hypothesis, we expect to observe a main effect of cloze probability in the N400 time interval: the N400 to the target verb is expected to be reduced in the high cloze condition relative to the low cloze condition regardless of whether the verb is preceded by an identical set of words.

Although these hypotheses do not make clear predictions regarding other potential effects of argument substitution, we independently expect that the target verbs may elicit a larger late positivity (P600) in the low cloze conditions than in the high cloze conditions as low cloze verbs may be considered less semantically plausible than high cloze verbs.

Methods

Participants

Twenty-four students (14 female, mean age = 21.8 years, range 18 - 27 years) from the University of Maryland, College Park participated in the current study. All participants were native speakers of English, 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 10 USD/hour for their participation. Data from three additional participants were excluded due to low accuracy on the judgment task (zero or negative d-prime in one or more of the conditions).

Materials

The experimental stimuli consisted of 120 sets of sentences with four versions each. All four versions in each set had the same target verb (the main verb in the embedded question). As in Experiment 1, the target was always preceded by the auxiliary "had", which disambiguated the structural assignment of the target word. A sample set of experiment items is shown in Table 7.

As in Experiment 1, the experimental materials were developed in stages. Since context was a within-items manipulation in the current experiment, twice as many argument substitution sentences were needed in this experiment as in Experiment 1. We adapted the argument substitution sentences used in Experiment 1 and created additional sentence frames for a separate cloze probability norming study. Sentence frames were created in sets of four, in which the fronted object was substituted in two ways. The different-words argument substitution manipulation was identical to the argument substitution manipulation in Experiment 1. The fronted object in one version was substituted by a different but discourse-compatible noun in the other version, while the rest of the sentence was held constant. This resulted in sentence pairs that differed by one word (e.g., "The tenant inquired which neighbor the landlord had ___" and "The tenant inquired which exterminator the landlord had ___"). Meanwhile, the same-words argument substitution manipulation changed one of the arguments of the target verb while holding the lexical contents of each sentence pair constant. The substituted fronted object in one version was used as the main clause subject in the other version (e.g., "The exterminator inquired which neighbor the landlord had ___" and "The neighbor inquired which exterminator the

landlord had ___"). Only common nouns (i.e., no proper names) were used in the main clause subject position, fronted object and embedded subject positions. A total of 106 sets of four sentence frames were created and divided into four lists, and each list was completed by 30 participants. The norming procedure was identical to the one used in Experiment 1.

We created 120 experimental items from the resulting database. Ideally, these items would consist of 120 unique sets of four sentence frames, each paired with one target verb. However, the feasibility of this approach was limited by the constraint that argument substitution must result in a clear cloze probability difference in both the same-words and the different-words conditions in each item. As a result, we created 120 items with 89 unique sets of sentence frames by pairing a subset of sentence frames with different target verbs to create different items. All items shared the property that the same target verb was used across all four conditions. Specifically, 58 items were created from 58 unique sets of sentence frames by pairing each set with one target verb, as in Experiment 1. Additionally, 62 items were created from 31 unique sets of sentence frames. This was done by pairing each set of sentence frames with two distinct verbs that showed opposite cloze probabilities. As illustrated in (1) and (2), two different target verbs ("evicted" and "hired") were paired with the same set of sentence frames to create two sets of experimental items. Crucially, the verb "evicted" has a higher cloze probability in "The tenant inquired which neighbor the landlord had ___" than in "The tenant inquired which exterminator the landlord had ___", while the opposite was true for the verb "hired."

(1) Sample context-target pairing:

(a) High cloze: "The tenant inquired which neighbor the landlord had evicted…"

(b) Low cloze: "The tenant inquired which exterminator the landlord had evicted…"

(2) Alternate context-target pairing:

(a) High cloze: "The tenant inquired which exterminator the landlord had hired…"

(b) Low cloze: "The tenant inquired which neighbor the landlord had hired…"

Argument substitution resulted in a clear cloze probability difference (at least 6%) in both the same-words and the different-words conditions in each item. The target verb had an average cloze probability of 22% in the high cloze conditions and less than 1% in the low cloze conditions (different-words: 0.4%; same-words: 0.8%). As in Experiment 1, in order to avoid end of sentence wrap up effects the sentences were extended beyond the critical verb with words that were identical across conditions within each item set.

The 120 sets of experimental sentences were distributed in four presentation lists, such that one member of each set appeared in each list. Importantly, since 31 sets of sentence frames were used in 62 items, care was taken to ensure that no two identical sentence frames appeared in the same list. In all presentation lists, items that originated from the same set of sentence frames always appeared in different blocks and they differed by their main clause subject, target verb as well as post-target words. In addition, each list also contained 120 filler sentences. The same 60 filler sentences from Experiment 1 were used for the control comparison. The other 60 fillers consisted of a separate cloze probability manipulation in mildly constraining sentences (cloze probability of high cloze words < 40%, mean = 24%) that are not discussed further here. In all filler sentences, each sentence context had a high cloze and a low cloze target word as a continuation; the high cloze target word in one item appeared as the low cloze target word (zero cloze) in another item. The overall high-to-low cloze ratio in each presentation list was 1:1. The 240 sentences in each list were presented in five blocks of 48 sentences each. The order of the blocks and the sentences within each block were randomized across participants.

Procedure

The experimental procedures were identical to those in Experiment 1.

EEG Recording

The EEG recording procedures were identical to those in Experiment 1.

ERP and Behavioral Data Analysis

The procedures for behavioral and ERP data analysis were identical to those in Experiment 1. Due to artifacts in the EEG data, a total of 14.4% of experimental trials, roughly equally distributed across conditions (13.9% - 15.0%), were excluded from the averaging procedure. All effects involving context or cloze probability are reported in the main text.

Results

Plausibility Judgments

As shown in Table 8, participants judged experimental sentences in the high cloze conditions to be plausible at a much higher rate than those in the low cloze conditions. Mixed-effects logistic regression revealed significant main effects of both factors (cloze: β = -.94, p(Wald) < .0001; context: β = .19, p(Wald) < .0001) and no interaction between them. Sentences in the high cloze conditions were judged 'plausible' around 40% more often than in the low cloze conditions; sentences in the different-words conditions were judged 'plausible' around 7% more often than in the same-words conditions. Further, the effect of cloze probability did not differ between same-words and different-words conditions. The non-zero values in the plausibility judgments in the low cloze conditions reflected the fact that low cloze sentences varied along a continuum of plausibility.

<Table 8 about here>

Event-related Potentials (experimental comparisons)

<Figure 4 about here>

<Figure 5 about here>

Figure 4 shows the grand average ERPs at the target word and the topographic distribution of the effects in the 300-500 ms and 700-900 ms intervals in both experimental and control comparisons in Experiment 2.

Figure 5 shows the grand average ERPs at the target word across all 9 ROIs in the different-words and same-words conditions. An N400 effect was observed in both the same-words and the different-words conditions: the N400 was smaller for high cloze than low cloze target verbs in both conditions. Following the N400 effect, low cloze target verbs elicited a larger positivity (P600) than high cloze target verbs in the different-words conditions but not in the same-words conditions.

These observations were confirmed by the statistical analyses. Results from the omnibus ANOVAs are presented in Table 9. The standardized effect size (Pearson's *r*) and 95% CI for the contrasts between the high and low cloze conditions at 4 midline central-posterior electrodes in the N400 and P600 time intervals are presented in Table 10. In the 300-500 ms interval, the omnibus ANOVA revealed a significant main effect of cloze probability and no interaction between cloze probability and context. The N400 was smaller for high cloze than low cloze targets in both the same-words and the different-words conditions. Further, the highly comparable effect sizes of the high vs. low cloze contrast in the same-words and different-words conditions show that the effect of cloze probability on the N400 did not differ between the same-words and different-words conditions. In the 700-900 ms interval, the omnibus ANOVA revealed a cloze probability × context interaction. Paired-sample t-tests within each level of context on ERPs averaged across all 18 electrodes revealed a significant effect of cloze probability in the different-words conditions ($t(23) = -2.15$; $p < .05$) but not in the same-words conditions ($|t| < 1$). ERPs were more positive across the scalp for low cloze

targets than high cloze targets in the different-words condition, but were not modulated by cloze probability in the same-words conditions.

<Table 9 about here>

<Table 10 about here>

Event-related Potentials (control comparison)

The grand average ERPs at the high and low cloze target words for the control comparison are shown in the lower part of Figure 4. As in Experiment 1, the control comparison showed that comprehenders also displayed clear N400 sensitivity to the standard cloze manipulation in the filler sentences in this experiment. The N400 was smaller for high cloze than low cloze words, and low cloze words elicited a larger late positivity than high cloze words.

These observations were supported by the statistical analyses. Results from the omnibus ANOVAs are presented in Table 5. The standardized effect size (Pearson's *r*) and 95% CI for the contrasts between the high and low cloze conditions at 4 midline central-posterior electrodes in the N400 and P600 time intervals are presented in Table 6. In the 300-500 ms interval the omnibus ANOVA revealed a main effect of cloze probability and a two-way cloze probability × anteriority interaction, due to the fact that the N400 effect was largest at posterior sites. In the 700-900 ms interval the omnibus ANOVA revealed a main effect of cloze probability and two-way interactions between cloze probability and each topographic factor, due to the fact that the late positivity was largest at central-posterior sites and slightly right-lateralized.

Discussion

In this experiment we asked whether comprehenders' initial predictions about the lexical semantic features of an upcoming verb are equally dependent on all preceding words in the sentence

(the Bag-of-words hypothesis), or whether it relies selectively on the verb's arguments (the Bag-of-arguments hypothesis). We find evidence for the latter. Crucially, we found that argument substitution elicited a significant N400 effect at the verb regardless of whether the sentence context contained an identical set of words.

The current results are not compatible with our formulation of the Bag-of-words hypothesis, according to which all preceding words (arguments or not) impact comprehenders' predictions in a uniform, bag-of-words fashion. Under this account, argument substitution should have no effects on the N400 at the verb as long as the verb is preceded by the same set of words in the sentence, as in the same-words conditions in the current experiment.

However, one might argue that these results may be compatible with an alternative formulation of the Bag-of-words hypothesis, according to which the effect of lexical priming diminishes over time, such that words that appeared more recently have a stronger impact than words that were further away. We will call this the "Bag-of-words + decay" hypothesis. According to this hypothesis, initial verb predictions are driven by the meaning of all preceding words (i.e., a bag of words), and the observation of an N400 effect in the same-words conditions is attributed to stronger lexical priming from words that are nearby (e.g., the fronted object and the embedded subject) than from those that are further away (e.g., the main clause subject). We tend to disprefer this hypothesis because we did not observe even a small effect of the main clause subject on the N400 to the verb. However, it is in principle possible that the main verb subject decayed so rapidly that it could have no facilitative effect on the processing of the embedded verb at all. Future work could address this alternative hypothesis by examining whether the N400 to the verb becomes sensitive to the main clause subject when the time elapsed between them is reduced.

We argue that the current results are best captured by the Bag-of-arguments hypothesis, according to which comprehenders quickly identify the arguments and initially use their lexical

meaning (but not their structural roles) to predict an upcoming verb. In the present experiment, argument substitution had resulted in a change in the verb's cloze probability and elicited an N400 effect in both the same-words and different-words conditions. Further, the size of the effect of argument substitution on the N400 at the verb was highly comparable across the same-words and different-words conditions (see Table 10). This suggests that the ease of accessing the verb in long-term semantic memory critically depends on the lexical meaning of that verb's arguments, which is in line with the Bag-of-arguments hypothesis.

Following the N400 response, we found that argument substitution elicited a P600 effect in the different-words conditions but not in the same-words conditions in the 700-900 ms time interval. Even though we did not have clear predictions regarding the effects of argument substitution on the P600 response, the presence of a P600 effect in one condition but not in the other was not predicted by any of the hypotheses discussed above. Since the target verb had a non-zero cloze probability in the low cloze conditions in some items, there was a considerable variability in the plausibility of sentences in the low cloze conditions in the current experiment. As evidenced by the plausibility judgment results, there was not a perfect correspondence between a target verb's cloze probability (high vs. low) and the sentence's plausibility (plausible vs. implausible), as some sentences in the low cloze conditions were judged as plausible and vice versa. As we reviewed in the Introduction, many studies have shown that the P600 can be modulated by a word's semantic congruity. Therefore, it is possible that sentences that were judged implausible elicited a larger P600 response than those that were judged plausible, but the imperfect correspondence between cloze probability and plausibility had resulted in an apparent interaction between cloze probability and context in the current experiment.

To examine this possibility, we analyzed the ERP data by grouping individual trials based on participants' plausibility judgments instead of the target verb's cloze probability. For each participant,

we computed average ERPs in trials that were judged as plausible and those that were judged as implausible. We conducted a 2 × 2 × 2 × 3 repeated measures ANOVA with context, plausibility, laterality and anteriority as within-participant factors on the by-participant average amplitude in the 700-900 ms time interval; the Greenhouse-Geisser correction was applied to univariate *F*-tests with more than one degree of freedom in the numerator. Although participants tended to have more "Yes" responses than "No" responses, there were similar rates of "Yes" responses across the same-words (64%) and different-words conditions (59%). This analysis revealed a main effect of plausibility ($F(1,23) = 12.46$, $p < .01$), and no interaction between plausibility and context (all Fs < 1). This showed that the P600 response at the target verb was larger in sentences that were ultimately judged as implausible compared to sentences that were ultimately judged as plausible, and that this effect did not differ between the same-words and different-words conditions. This is consistent with Van Petten and Luka's (2012) proposal that the P600 response is modulated by a word's semantic congruity. However, as participants showed a similar pattern of plausibility judgments across the same-words and different-words conditions (Table 8), this still cannot explain the presence of a P600 effect in the different-words conditions, but not the same-words conditions.

General Discussion

In the current study we took a first step towards isolating different sources of contextual information and compared how they impact predictive computations in real-time language comprehension. We focused on the impact of the lexical meaning and structural roles of preverbal arguments on comprehenders' lexical semantic predictions about an upcoming verb in verb-final embedded questions in English. We manipulated argument role information by reversing the order of the preverbal arguments, and changed the lexical semantic information supplied by the arguments by substituting one of the preverbal arguments. Using the amplitude of the N400 at the verb as an index

of the ease of accessing the verb in long-term semantic memory, we inferred the extent to which a verb (or its lexical semantic features) is pre-activated by each of these two sources of information by examining the effect of argument role reversal and argument substitution on N400 amplitudes at the verb.

In Experiment 1 we found that, even though argument substitution and role reversal affected the verb's offline predictability (cloze probability) equally, only argument substitution elicited an N400 effect at the verb. In Experiment 2 we found that argument substitution elicited an N400 effect at the verb even when the lexical items in the sentence context were matched across conditions.

The N400's robust sensitivity to cloze probability differences that resulted from argument substitution (Experiments 1 and 2) stood in contrast with its complete insensitivity to cloze probability differences that resulted from argument role reversals (Experiment 1). This contrast was sharpened by the extent to which argument role reversal and argument substitution were matched in various ways. First, both manipulations were implemented in sentences with the same syntactic structure (i.e., an embedded object question), such that the linear and structural distance between the target verb and its arguments were held constant. In addition, we carefully matched the effects of argument role reversal and argument substitution on the target verb's cloze probability in the experimental sentences. Further, neither the argument role reversal manipulation in Experiment 1 nor the same-words argument substitution manipulation in Experiment 2 resulted in any lexical differences between the high and low cloze conditions. In the argument role reversal conditions in Experiment 1, the only difference between the high and low cloze conditions was the order (and hence the roles) of the arguments of the embedded verb; in the same-words argument substitution conditions in Experiment 2, the only difference between the high and low cloze conditions was the order of a different pair of nouns in the sentence context: the fronted object in the embedded question in one condition became the main clause subject in the other condition and vice versa.

We propose that the N400's sensitivity to cloze differences resulting from argument substitution, even when the sentence contexts contained identical sets of words, suggests that comprehenders quickly distinguish the arguments from other words in the context and pre-activate verbs that are thematically related to the arguments. In contrast, the N400's insensitivity to cloze differences resulting from argument role reversal suggests that, by the time it appeared in the input, the target verb was not differentially pre-activated based on information about its arguments' structural roles. Note, however, since the arguments are the only nouns in the same clause as the target verb in the current experimental materials, these results cannot rule out the possibility that comprehenders use all nouns in the same clause as the verb, e.g., non-arguments as well as arguments, to predict an upcoming verb. Future research will need to explore this possibility (also see Stroud 2008, Chapter 3, for broader discussion of comprehenders' sensitivity to the argument status of nouns).

A 'bag-of-arguments' mechanism for initial verb predictions

We propose that verb predictions in a verb-final clause initially rely on a 'bag-of-arguments' mechanism, in which comprehenders quickly distinguish the arguments from other words in the context and pre-activate verbs that are thematically related to the arguments, before taking into consideration the structural roles of the arguments. Naturally this raises a question: why might the preverbal arguments' lexical meaning have a more immediate impact on verb predictions than their structural roles?

One possible explanation is that the sentences are initially misparsed and have to be reparsed. If comprehenders initially analyze the *wh*-phrase (e.g., "which customer") as a displaced subject, then as the next words (e.g., "the waitress") appear and signal that the subject position has already been filled, comprehenders would have to revise their initial parse, which may in turn delay the impact of argument role information on verb predictions. However, since all experimental sentences contained

an embedded object question and none of our filler sentences contained an embedded subject question, we consider it unlikely that participants would have consistently misparsed the sentences throughout the experiment. Further, studies in other languages have shown that argument role reversals fail to elicit an N400 effect even when the sentence structure is unambiguous (e.g., Kolk et al., 2003; Oishi & Sakamoto, 2010; Chow & Phillips, 2013), and hence we argue that the delayed impact of argument role information should not be attributed to misparsing of the sentence.

Alternatively, argument role information may have a delayed impact on verb predictions because information about the arguments' likely thematic roles becomes available to comprehenders later in time than their lexical meaning. The meaning of an argument may pre-activate thematically related verbs as soon as it has been retrieved from the lexicon. Meanwhile, in order to use the structural roles of the arguments for verb prediction, comprehenders must first map the arguments' structural positions to likely thematic roles. Therefore, even if comprehenders can quickly compute the structural position of the arguments in the sentence (as shown by their ability to readily distinguish arguments from other words in the sentence), the arguments' structural roles may have a delayed impact on comprehenders' verb predictions because the computation of an argument's likely thematic roles based on its structural position may take more time than the retrieval of a words' meaning from the lexicon.

Last but not least, it is possible that even the likely thematic roles are computed very quickly, but that it takes time to use this information to access suitable verb candidates in memory. For example, a search in long-term memory for events that involve waitresses (regardless of thematic roles) might be rather fast, but it might take longer to search memory for events that involve waitresses specifically as agents. Alternatively, one might go about searching for waitress-as-agent events in memory by first retrieving events that involve waitresses and then filtering out the ones in which the waitress is not an

agent. In this case, the delayed impact of arguments' structural roles on verb prediction may reflect how event knowledge is queried and retrieved from long-term memory.

"Fast" vs. "slow" predictions

Our findings are compatible with various findings that certain types of linguistic predictions are computed rather quickly, but suggest that it is important to distinguish predictions based on what is predicted (e.g., word category vs. lexical semantic information vs. phonological/orthographic features) and what cues are used to generate those predictions. Here we show that the lexical semantic features of the arguments are rapidly used to pre-activate likely upcoming verbs (or their lexical semantic features). In other work, Dikker et al. (2009) observed an increased early M100 response localized to visual cortex to a noun when it was preceded by an adverb compared to when it was preceded by an adjective, and argued that comprehenders can quickly generate expectations about an upcoming word's syntactic category, which in turn affects their expectations about the visual form of the upcoming word. This proposal suggests that in the current study, upon processing the arguments, comprehenders may quickly anticipate a verb to appear next, even when their verb predictions may not be immediately sensitive to all information about the arguments.

Similarly, Omaki and colleagues' Hyper-active Gap Filling account proposes that, upon encountering a displaced object (a "filler"), comprehenders posit a post-verbal gap site without waiting for information about the transitivity of the verb (Omaki, Lau, Davidson White, Dakan, Apple & Phillips, 2015). Under this view, upon processing the arguments in the present study, comprehenders may quickly predict a transitive verb to appear next, while detailed information about the arguments may only impact their specific predictions about the meaning and argument structure of an upcoming verb at a later time point. Future work may incorporate the current results with this broader pattern of findings to develop a detailed model of what kinds of verbal information are predicted when.

Offline vs. online measures of prediction

Along with our previous findings in Mandarin Chinese, the current results demonstrate a clear discrepancy between offline measures of a word's predictability and comprehenders' online predictions. Offline predictability measures such as those obtained in language corpora and offline cloze tasks provide useful information about the relations between a word and its preceding context. However, in order to examine how context impacts predictions in real time, we must carefully consider when different sources of contextual information arise and how they might impact predictive computations in real time.

The P600 and (im)plausibility

We observed a P600 effect in a few comparisons in the current study. First, the standard cloze probability manipulation in the filler sentences elicited a P600 effect following the N400 effect in both experiments. In Experiment 1 we also observed a P600 effect in both the argument role reversal and argument substitution conditions. As participants judged sentences in the low cloze conditions to be less plausible than those in the high cloze conditions across all these comparisons, the presence of a P600 effect in these comparisons is consistent with previous reports that the P600 is modulated by the semantic congruity of a word (for a review see Van Petten & Luka, 2012). Note, however, that semantic incongruities do not always elicit a P600 effect. Van Petten and Luka's (2012) meta-analysis revealed that semantic incongruity elicited a P600 effect in one-third of the 64 published comparisons surveyed. The probabilistic nature of the P600's sensitivity to semantic congruity was observed in Experiment 2, where argument substitution elicited a P600 effect in the different-words conditions but not in the same-words conditions. Even though further analyses showed that the P600 at the target verb was larger in sentences that were ultimately judged as implausible compared to sentences that were judged as plausible, we do not yet have an account for the contrast between the same-words

and different-words conditions. These results suggest that even the same group of comprehenders may show non-uniform P600 sensitivity to semantic congruity. Future research will be needed to examine factors that may contribute to the P600's probabilistic sensitivity to semantic congruity.

Conclusion

In the current study we examined how the lexical meaning and the structural roles of preverbal arguments impact comprehenders' verb predictions in real time. We found electrophysiological evidence that comprehenders immediately use the lexical meaning of the arguments, but not their structural roles, to pre-activate likely upcoming verbs (or their lexical semantic features). We also saw that the lexical meaning of non-arguments in the sentence did not impact verb predictions to the same extent. We propose that verb prediction immediately and specifically relies on the arguments in a clause, while initially failing to take into consideration the information provided by the thematic roles of those arguments.

Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

Acknowledgements

This work was supported by the National Science Foundation under Grant BCS-0848554 to Colin Phillips. We thank Glynis MacMillan, Shefali Shah, and Erin Mahoney for invaluable help in materials creation and data collection.

References

Altmann, G.T.M., Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition,* 73, 247–264.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280-289.

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience, 29*, 8447-8451.

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***,** 489-505.

Boland, J.E., Tanenhaus, M.K., Garnsey, S.M., and Carlson, G. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, 34, 774-806.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127-143.

Chow, W.Y., & Phillips, C. (2013). No semantic illusion in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain Research, 1506, 76–93.*

Chow, W.Y., Lau, E., Wang, S., & Phillips, C. (Submitted). Timing is everything: the temporal dynamics of word prediction.

Deacon, D., Hewitt, S., Yang, C.M., & Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: Evidence that the N400 does not reflect a post-lexical process. *Cognitive Brain Research*, 9, 137-146.

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

Demberg, V., Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition,* 101, 193–210.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41(6), 391-407.

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to Syntax in Visual Cortex. *Cognition*, 110, 293-321.

Federmeier, K.D., Wlotko, E.W., De Ochoa-Dewald, E., Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research,* 1146: 75-84.

Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

Federmeier, K.D., & Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *Journal of Memory and Language*, 41, 469-495.

Garnsey, S.M., Tanenhaus, M.K., & Chapman, R.M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, 18, 51-60.

Gerlanc, D., Kirby, K.N. (2013). bootES (Version 1.0.1). Retrieved from http://cran.r-project.org/web/packages/bootES/index.html

Gibson, E., Stearns, L., Bergen, L., Eddy, M., & Fedorenko, E. (2013). The P600 indexes rational error correction within a noisy-channel model of human communication. Talk presented at the 26th annual CUNY Human Sentence Processing Conference, Columbia, SC.

Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438-441.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL-2001* (pp. 159–166). Stroudsburg, PA: Association for Computational Linguistics.

Hickok, G., Canseco-Gonzalez, E., Zurif, E., & Grimshaw, J. (1992). Modularity in locating wh-gaps. *Journal of Psycholinguistic Research,* 21 (6), 545-561

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, 59-73.

Jaeger, T.F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language,* 59 (4), 434—446.

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

Kim, A. & Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language*, 118, 15-22.

Kolk, H.H.J., Chwilla, D.J., van Herten, M., & Oor, P. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain and Language*, 85, 1-36.

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: 222. doi: 10.3389/fpsyg.2010.00222

Kukona, A., Fang, S., Aicher, K.A., Chen, H., & Magnuson, J.S. (2011). The time course of anticipatory constraint integration. *Cognition,* 119, 23-42.

Kutas, M., & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4, 463-470.

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

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.

Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

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

Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211–240.

Landauer, T.K., Foltz, P.W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259-284.

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

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition,* 106: 1126–1177.

Nieuwland, M.S., & Kuperberg, G.R. (2008). When the truth isn't too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19, 1213-1218.

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*, 9, 97–113.

Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*. 6:384. doi: 10.3389/fpsyg.2015.00384

Osterhout, L., & Holcomb, P.J. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.

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.

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 (4), 426-448.

Phillips, C., Wagers, M.W., & Lau, E.F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (ed.), *Experiments at the Interfaces*, Syntax & Semantics, vol. 37, pp. 153-186. Bingley, UK: Emerald Publications.

Smith, N.J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition,* 128, 302–319.

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

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

Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

Van Herten, M., Chwilla, D.J., & Kolk, H.H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience* 18, 1181-1197.

Van Herten, M., Kolk, H.H., & Chwilla, D.J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research,* 22, 241-255.

Van Petten, C., Luka, B.J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology,* 83, 176–190.

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, 1272-1288.

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*, 356-366.

Wolpert, D.M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences,* 1, 209-216.

Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

Tables with captions

Table 1. Experimental conditions and sample materials in Experiment 1.

Argument role reversal

The target verb's cloze probability was manipulated by reversing the order of the arguments in the embedded question.

	Sample Material	Target cloze mean (sd)
High cloze	The restaurant owner forgot which customer the waitress had **served** during dinner yesterday.	.25 (.09)
Low cloze	The restaurant owner forgot which waitress the customer had **served** during dinner yesterday.	0 (0)

Argument substitution

The target verb's cloze probability was manipulated by substituting the fronted object in the embedded question.

	Sample Material	Target cloze mean (sd)
High cloze	The superintendent overheard which tenant the landlord had **evicted** at the end of May.	.28 (.14)
Low cloze	The superintendent overheard which realtor the landlord had **evicted** at the end of May.	0 (0)

Table 2. Summary of plausibility judgment in Experiment 1.

Experiment 1	Mean "Plausible" judgments % (SE)
Argument role reversal, high cloze	85.4 (2.7)
Argument role reversal, low cloze	23.8 (3.1)
Argument substitution, high cloze	90.3 (2.2)
Argument substitution, low cloze	31.1 (2.9)

Table 3. ANOVA F-values at the target verb in Experiment 1.

	df	300-500ms	700-900ms
Omnibus ANOVA			
context	1,23	1.71	< 1
cloze	1,23	5.37*	5.57*
context × cloze	1,23	4.88*	< 1
context × ant	2,46	3.75*	< 1
cloze × ant	2,46	< 1	1.42
context × lat	2,46	< 1	< 1
cloze × lat	2,46	6.54**	1.98
context × cloze × ant	2,46	< 1	1.1
context × cloze × lat	2,46	< 1	< 1
context × ant × lat	4,92	2.3^	< 1
cloze × ant × lat	4,92	2.57^	2.22^
context × cloze × ant × lat	4,92	1.49	1.15

Topographic factors: ant = anterioriy; lat = laterality.
** $p < .01$
* $p < .05$
^ $.05 < p < .1$

Table 4. Standardized effect size (Pearson's *r*) and 95% confidence interval for the contrast between the high and low cloze experimental conditions using the mean amplitude at 4 midline central-posterior electrodes (CZ, CPZ, PZ, OZ) in the N400 and P600 time intervals in Experiments 1.

	300-500ms		700-900ms	
Effect of cloze probability	Effect size	95% CI	Effect size	95% CI
Argument role reversal	-0.08	[-0.48, 0.33]	0.44	[0.10, 0.65]
Argument substitution	-0.51	[-0.71, -0.16]	0.23	[-0.19, 0.57]

Table 5. ANOVA F-values at the target word in the control comparison in Experiments 1 and 2.

	df	300-500ms	700-900ms
Experiment 1			
cloze	1,23	21.84**	12.12**
cloze × ant	2,46	5.71*	< 1
cloze × lat	2,46	3.77*	1.97
cloze × ant × lat	4,92	3.78*	1.3
Experiment 2			
cloze	1,23	12.24**	12.78**
cloze × ant	2,46	6.21**	4.78*
cloze × lat	2,46	< 1	4.82*
cloze × ant × lat	4,92	1.36	< 1

Factors: cloze = cloze probability; ant = anteriority; lat = laterality.

** $p < .01$

* $p < .05$

^ $.05 < p < .1$

Table 6. Standardized effect size (Pearson's *r*) and 95% confidence interval for the control comparison (low cloze minus high cloze) using the mean amplitude at 4 midline central-posterior electrodes (CZ, CPZ, PZ, OZ) in the N400 and P600 time intervals in Experiments 1 and 2.

	300-500ms		700-900ms	
Effect of cloze probability	Effect size	95% CI	Effect size	95% CI
Control comparison (Exp 1)	-0.73	[-0.86, -0.47]	0.52	[0.03, 0.74]
Control comparison (Exp 2)	-0.61	[-0.78, -0.32]	0.59	[0.33, 0.76]

Table 7. Experimental conditions and sample materials in Experiment 2.

Different words argument substitution (Replication of Experiment 1)
The fronted object in the embedded question was substituted; the sentence frames were otherwise identical between conditions.

	Sample Material	Target cloze mean (sd)
High cloze	The tenant inquired which neighbor the landlord had **evicted** from the apartment complex.	.22 (.16)
Low cloze	The tenant inquired which exterminator the landlord had **evicted** from the apartment complex.	.004 (.01)

Same words argument substitution
The fronted object in the embedded question was substituted; the substituted argument in one condition was used as the main clause subject in the other condition to create sentence frames that contained the same set of words.

	Sample Material	Target cloze mean (sd)
High cloze	The exterminator inquired which neighbor the landlord had **evicted** from the apartment complex.	.22 (.14)
Low cloze	The neighbor inquired which exterminator the landlord had **evicted** from the apartment complex.	.008 (.02)

Table 8. Summary of plausibility judgments in Experiment 2.

Experiment 2	Mean "Plausible" judgments % (SE)
Different words, high cloze	83.1 (2.2)
Different words, low cloze	46.7 (3.7)
Same words, high cloze	77.2 (2.4)
Same words, low cloze	39.2 (3.6)

Table 9. ANOVA F-values at the target verb in Experiment 2.

	df	300-500ms	700-900ms
Omnibus ANOVA			
context	1,23	< 1	1.05
cloze	1,23	5.5*	1.1
context × cloze	1,23	< 1	6.12*
context × ant	2,46	< 1	< 1
cloze × ant	2,46	3.63^	< 1
context × lat	2,46	1.13	< 1
cloze × lat	2,46	< 1	2.4
context × cloze × ant	2,46	1.96	1.07
context × cloze × lat	2,46	< 1	< 1
context × ant × lat	4,92	< 1	< 1
cloze × ant × lat	4,92	1.15	< 1
context × cloze × ant × lat	4,92	< 1	< 1

Topographic factors: ant = anteriority; lat = laterality.
** $p < .01$
* $p < .05$
^ $.05 < p < .1$

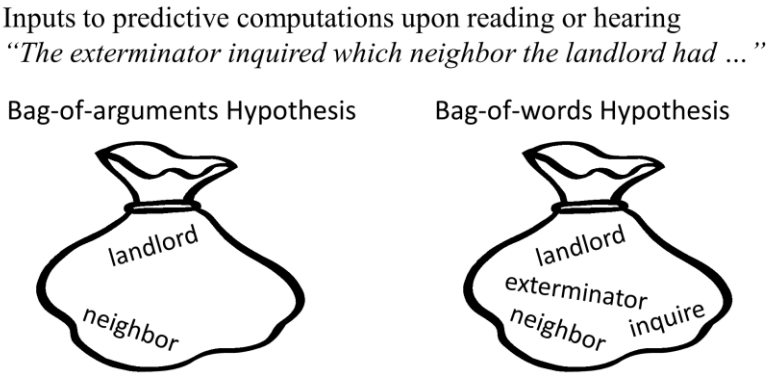
Table 10. Standardized effect size (Pearson's *r*) and 95% confidence interval for the contrast between the high and low cloze experimental conditions using the mean amplitude at 4 midline central-posterior electrodes (CZ, CPZ, PZ, OZ) in the N400 and P600 time intervals in Experiment 2.

	300-500ms		700-900ms	
Effect of cloze probability	Effect size	95% CI	Effect size	95% CI
Different words	-0.36	[-0.67, 0.05]	0.37	[-0.03, 0.64]
Same words	-0.45	[-0.65, -0.02]	-0.23	[-0.52, 0.21]

Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

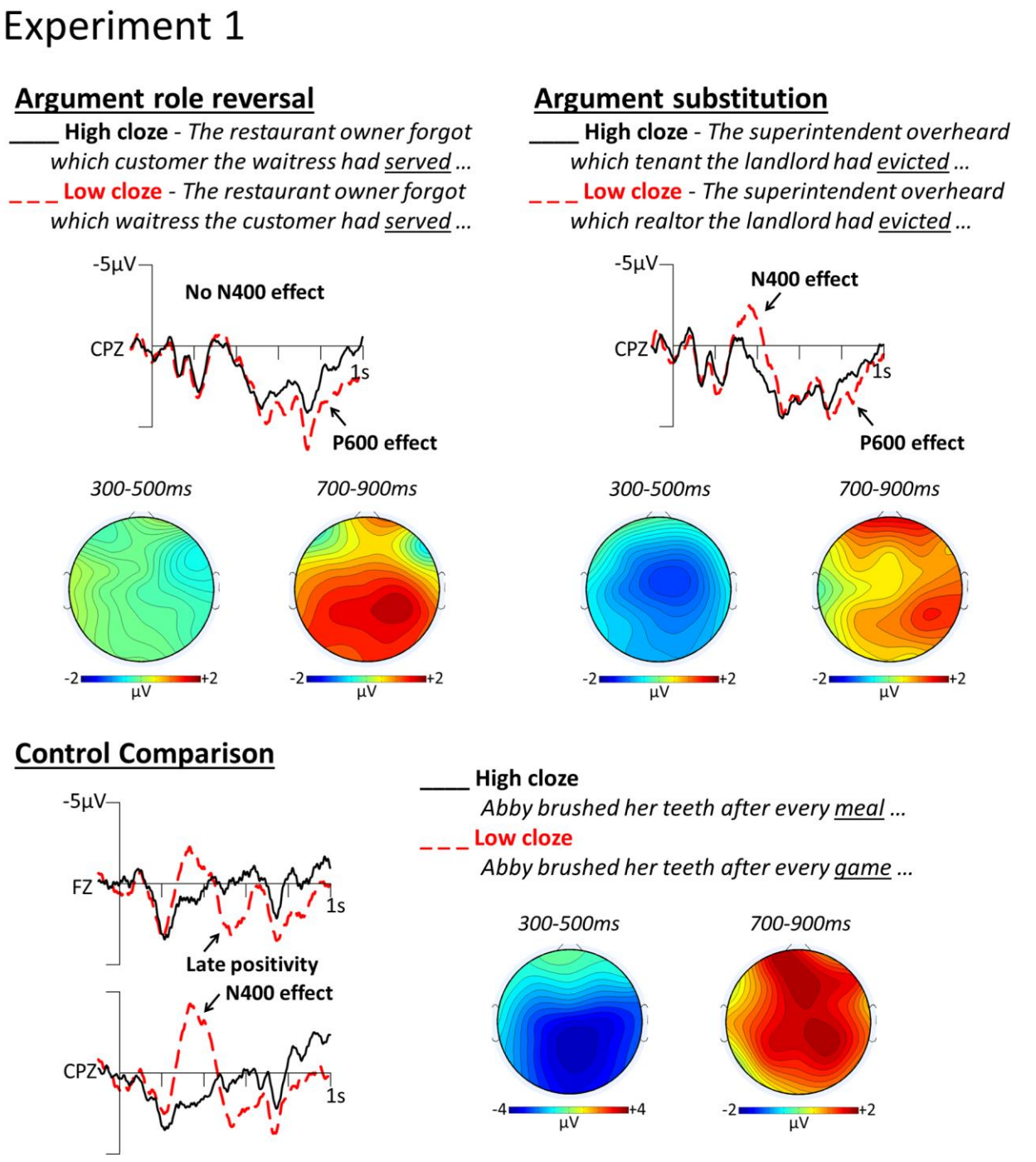
Figures with captions

Figure 1. A schematic representation of the inputs to predictive computations under the Bag-of-arguments hypothesis (left) and the Bag-of-words hypothesis (right).



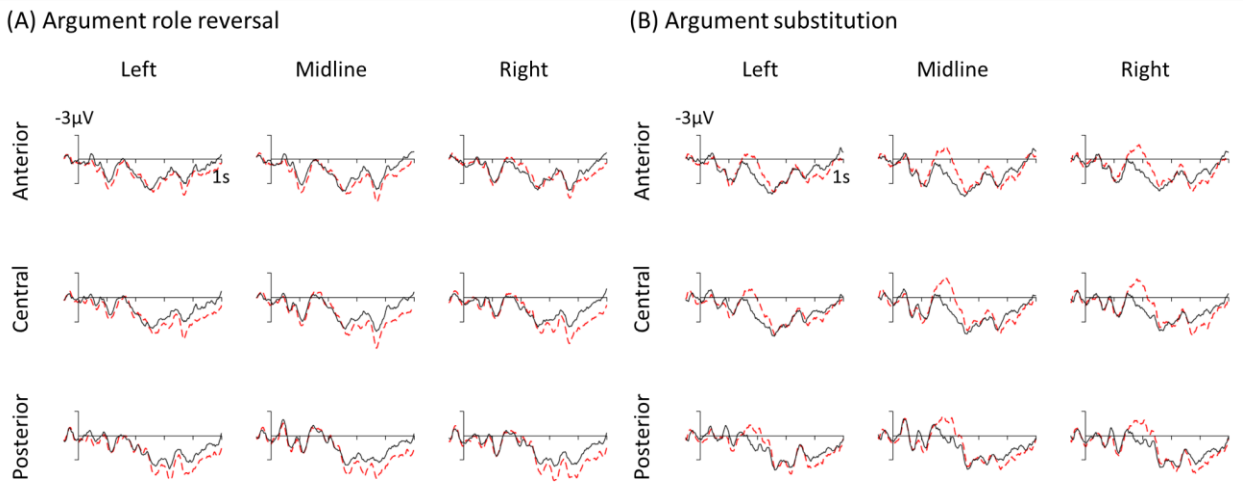
Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

Figure 2. *Top*: Grand average ERPs at centro-posterior electrode CPZ and topographic distribution of ERP effects in the 300-500 ms and 700-900 ms intervals (low cloze minus high cloze) in the argument role reversal (left) and argument substitution (right) conditions in Experiment 1. *Bottom*: Grand average ERPs at frontal electrode FZ and at centro-posterior electrode CPZ and topographic distribution of ERP effects in the 300-500 ms and 700-900 ms intervals (low cloze minus high cloze) in the control comparison in Experiment 1.



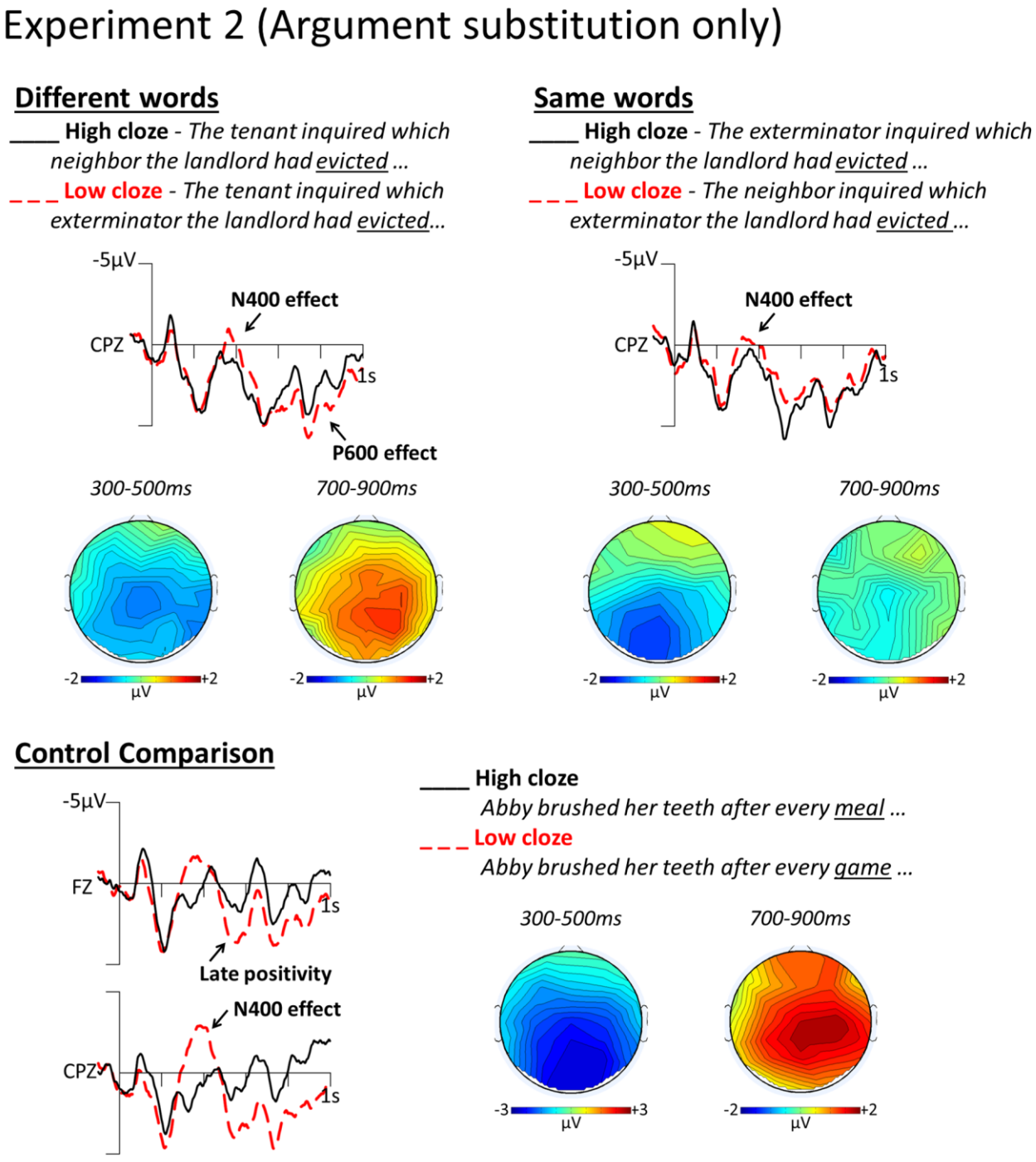
Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

Figure 3. Grand average ERPs across nine regions of interest at the high cloze (black solid line) and low cloze (red dashed line) target verb in the argument role reversal (left) and argument substitution (right) conditions in Experiment 1.



Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

Figure 4. *Top*: Grand average ERPs at centro-parietal electrode CPZ and topographic distribution of ERP effects in the 300-500 ms and 700-900 ms intervals (low cloze minus high cloze) in the argument role reversal (left) and argument substitution (right) conditions in Experiment 1. *Bottom*: Grand average ERPs at frontal electrode FZ and at centro-posterior electrode CPZ and topographic distribution of ERP effects in the 300-500 ms and 700-900 ms intervals (low cloze minus high cloze) in the control comparison in Experiment 2.



Chow W-Y, Smith C, Lau E, Phillips C (2015). A 'bag-of-arguments' mechanism for initial verb predictions. Language, Cognition and Neuroscience, 1-20. DOI: 10.1080/23273798.2015.1066832

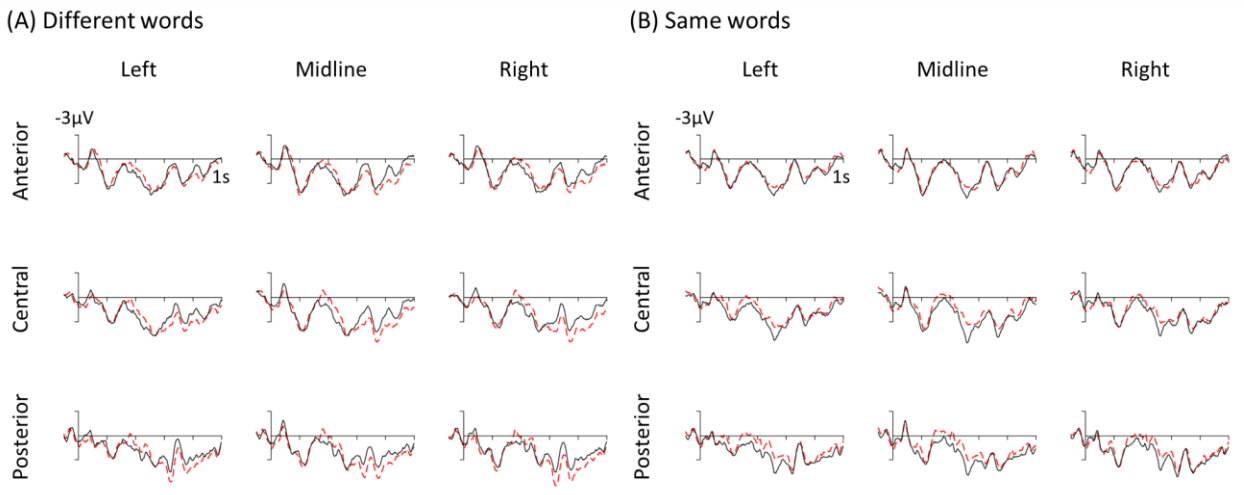



Figure 5. Grand average ERPs across nine regions of interest at the high cloze (black solid line) and low cloze (red dashed line) target verb in the different-words (left) and same-words (right) conditions in Experiment 2.