Psychophysiology, 00 (2015), 00–00. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/psyp.12515

Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials

BRENNAN R. PAYNE,[a,b] CHIA-LIN LEE,[d] AND KARA D. FEDERMEIER[a,b,c]

[a]Department of Psychology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[b]The Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[c]The Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[d]Graduate Institute of Linguistics, Department of Psychology, Graduate Institute of Brain and Mind Sciences, and Neurobiology and Cognitive Neuroscience Center National Taiwan University, Taipei, Taiwan

Abstract

The amplitude of the N400—an event-related potential (ERP) component linked to meaning processing and initial access to semantic memory—is inversely related to the incremental buildup of semantic context over the course of a sentence. We revisited the nature and scope of this incremental context effect, adopting a word-level linear mixed-effects modeling approach, with the goal of probing the continuous and incremental effects of semantic and syntactic context on multiple aspects of lexical processing during sentence comprehension (i.e., effects of word frequency and orthographic neighborhood). First, we replicated the classic word-position effect at the single-word level: Open-class words showed reductions in N400 amplitude with increasing word position in semantically congruent sentences only. Importantly, we found that accruing sentence context had separable influences on the effects of frequency and neighborhood on the N400. Word frequency effects were reduced with accumulating semantic context. However, orthographic neighborhood was unaffected by accumulating context, showing robust effects on the N400 across all words, even within congruent sentences. Additionally, we found that N400 amplitudes to closed-class words were reduced with incrementally constraining syntactic context in sentences that provided only syntactic constraints. Taken together, our findings indicate that modeling word-level variability in ERPs reveals mechanisms by which different sources of information simultaneously contribute to the unfolding neural dynamics of comprehension.

Descriptors: Event-related potentials (ERPs), Sentence comprehension, Linear mixed-effects model, N400, Lexical processing

The mechanisms underlying the comprehension of language are complex, involving a highly distributed set of neural networks brought online to ultimately form message-level meaning representations from sensory input. In particular, the comprehension of multiword text or utterances involves the continuous and incremental online mapping of incoming sensory stimuli onto incomplete semantic representations. Although a substantial behavioral literature in psycholinguistics has recently amassed that supports immediate and incremental online processing in sentence comprehension (Altmann & Kamide, 2007; Kamide, 2008; Rayner, 2009), evidence from event-related brain potentials (ERPs) for the incremental formation of semantic representations has existed since the 1980s (Kutas, Van Petten, & Besson, 1988; reviewed in Kutas & Van Petten, 1994). One oft-cited line of evidence supporting incremental semantic processing is that for open-class words, the amplitude of the N400—an ERP component linked to meaning processing and initial access to semantic memory (Kutas & Federmeier, 2000, 2011)—is inversely related to the buildup of semantic context over the course of a sentence (Van Petten & Kutas, 1990, 1991). This finding suggests that semantic information in the message-level representation builds incrementally with accruing context, thus easing the semantic access of meaningful later-occurring words.

In the current study, we revisited the nature of the incremental buildup of sentential semantic and syntactic context on the N400, adopting a flexible, item-level analysis via linear mixed-effects modeling (LMM). This approach allows for a fine-grained and continuous treatment of word-by-word variation on the event-related EEG in individual subjects, revealing single-item level influences on the N400. Importantly, our goal in the current study was to use the flexibility afforded by LMM of word-level event-related EEG to examine the degree to which accruing sentential context modulates multiple aspects of lexico-semantic word processing, as indexed by the N400.

The N400 is part of a default response to any potentially meaningful stimulus and is broadly sensitive to a whole host of factors that impact semantic processing (Kutas & Federmeier, 2000, 2011). In language processing, the N400 shows graded modulation

This research was supported by a James S. McDonnell Foundation Scholar Award and the National Institute on Aging (Grant AG026308) to K. D. Federmeier. We would like to thank the anonymous reviewers and the members of the UIUC language comprehension joint lab meeting for their helpful discussions and comments on earlier drafts of this article.

Address correspondence to: Brennan R. Payne, Beckman Institute, University of Illinois at Urbana-Champaign, 405 North Mathews Avenue, Urbana, IL 61801, USA. E-mail: payne12@illinois.edu

PSYCHOPHYSIOLOGY

based on the degree to which a stimulus is congruent with its prior semantic context (as operationalized, for example, by cloze probability; Kutas & Hillyard, 1984). Modulation of the N400 occurs continuously and cumulatively across multiple words within a sentence, even in the absence of explicit experimental manipulations or task demands. Van Petten, Kutas, and colleagues (Van Petten & Kutas, 1990, 1991; Van Petten, 1993) have shown that N400 amplitudes to open-class words are reduced with increasing ordinal word position within a congruent sentence, suggesting that the N400 is sensitive to the incremental buildup of semantic context (see also Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Halgren et al., 2002).

In the absence of higher-order discourse constraints (cf. Van Petten, 1995), a reader begins a sentence without message-level semantic information. However, with increasing progress into a sentence, a conceptual representation is incrementally built, reducing the demands on semantic access for subsequent words and, in some cases, also allowing the comprehender to anticipate and preactivate semantic features of likely upcoming words (cf. Federmeier, 2007; Kutas & Van Petten, 1994). Together, accumulated context-related semantic activation and increased predictability result in a reduction of N400 amplitude with increasing intrasentential word position. Van Petten and Kutas (1991) showed that the word-position effect is specific to open-class words in congruent sentences. Such word-position effects are not seen in randomly ordered words, or in so-called syntactic prose, wherein syntactic structure is maintained without a coherent message-level semantic interpretation (e.g., *The infuriated water grabbed the justified dream*; Marslen-Wilson & Tyler, 1980). They argued that cumulative semantic context effects on the N400 are global, building up over the course of an entire sentence.

Importantly, Van Petten and Kutas found that the influence of lexical properties of a word interacts with accumulating message-level semantic constraints. At the beginning of congruent sentences, for example, word frequency effects are robust, such that more frequent words show a smaller N400 compared to less frequent words. However, this word frequency effect is reduced as the sentence context accumulates (Van Petten & Kutas, 1990, 1991), suggesting that over the course of a sentence, semantic contextual constraints supersede the influence of lexical frequency on semantic processing. It has been argued that the sensitivity of the N400 to word frequency out of context reflects "baseline" (albeit task dependent, e.g., Fisher-Baum, Dickson, & Federmeier, 2014) semantic activity levels, with smaller levels for more frequent words. As a meaningful context builds, these baseline levels are overridden by higher-order semantic constraints (see Federmeier & Laszlo, 2009). Importantly, Van Petten and Kutas (1991) did not find such frequency by context interactions in syntactic prose and random sentences, indicating that this effect was due to an accumulating congruent message-level semantic representation.

More recently, studies of visual word recognition in and out of sentence contexts (Holcomb, Grainger, & O'Rourke, 2002; Laszlo & Federmeier, 2008, 2009, 2011, 2014; Van Petten, 2014; Vergara-Martínez & Swaab, 2012) have examined the influence of a word's orthographic neighborhood (the number and features of orthographically similar strings; Coltheart, Davelaar, Jonasson, & Besner, 1977; Yarkoni, Balota & Yap, 2008) on semantic processing. These studies report that N400 amplitudes are larger (more negative) for words with more orthographic neighbors. These findings suggest that information associated with highly orthographically similar items is initially activated in parallel with a presented word, such that words with many orthographic neighbors engender more activation in the semantic system.

Whereas frequency effects show clear modulation as a function of sentence context (Dambacher et al., 2006; Halgren et al., 2002; Van Petten & Kutas, 1990, 1991), it is not clear if orthographic neighborhood effects show the same sensitivity to accumulating sentence constraint. For example, in some work, orthographic neighborhood effects appear to be robust for strings at the end of highly constraining and congruent sentences (Laszlo & Federmeier, 2008). Federmeier and Laszlo (2009) have argued that divergent effects of context on the impact of frequency and orthographic neighborhood would be expected given the aspects of semantic memory use indexed by these lexical variables. As described earlier, frequency effects are argued to reflect transient and malleable activation states in semantic memory, initially arising as baseline activation states, but reducing with increasing message-level context. In contrast, neighborhood effects are argued to reflect intrinsic structural organization within the semantic system. To the extent that visual access to the semantic system is organized to some degree by similarity among orthographic inputs, effects of neighborhood would be expected to persist even in the presence of strongly constraining sentence contexts. However, other work suggests that sentence context may modulate orthographic similarity effects in visual word recognition. In a study by Molinaro, Conrad, Barber, and Carreiras (2010), words with high- or low-frequency orthographic neighbors completed sentences that were strongly or weakly contextually constraining. They found that orthographic neighbor frequency only modulated the N400 to words embedded in weakly constraining contexts. When sentence-final words were presented in strongly constraining contexts, there was no effect of neighbor frequency. The authors argued that the supportive sentence context preactivates specific target word forms, resulting in less competition from orthographically similar words. Thus, the literature shows mixed results with respect to whether orthographic similarity is modulated by sentential constraint.

To our knowledge, no study has systematically investigated the simultaneous impact of multiple lexical influences, such as orthographic neighborhood and word frequency, on semantic processing as they unfold over accumulating sentence contexts. ERP studies have not investigated such effects in part because traditional analyses preclude the simultaneous consideration of multiple continuous item-level covariates. For example, in studies examining incremental context effects on the N400, word position has been treated as a factor that is created by binning position into discrete (and sometimes uneven) levels (e.g., word 2, words 3–4, words 5+). Similarly, words vary continuously in their lexical attributes (frequency, neighborhood size, length, concreteness, etc.), but ERP studies typically compare items at extreme values of these distributions. The act of dichotomizing/discretizing continuous variables has long been known to reduce power and can result in spurious findings in certain cases (e.g., MacCallum, Zhang, Preacher, & Rucker, 2002). Although this is widely known, the practice of signal averaging across many items has often been viewed as necessary in ERP data analyses in order to identify components and improve the signal-to-noise ratio. However, a growing number of studies have demonstrated that item-level analyses of EEG/ERPs can result in reliable effects (Dambacher et al., 2006; Delorme, Miyakoshi, Jung, & Makeig, 2014; Frank, Otten, Galli, & Vigliocco, 2015; Gaspar, Rousselet, & Pernet, 2011; Laszlo & Federmeier, 2014; Tremblay & Newman, 2015; Van Petten, 2014), challenging the assumption that the signal-to-noise ratio may be too low to detect ERP effects at the level of single items.

Thus, an aim of the current study was to adopt an individual item–level analysis, utilizing LMM to examine word-to-word variation in the N400. The use of (generalized) linear mixed-effects models (also known as hierarchical linear models, multilevel models, or variance components models) has been prevalent in social science, biology, education, and behavioral research for some time (e.g., Singer, 1998; Snijders & Bosker, 2011). Recently, these modeling techniques have begun to gain ground in psycholinguistics, cognitive psychology, and cognitive neuroscience research as a tool to accommodate statistical dependency that arises from the kinds of nested and hierarchically structured data that are common in these fields (Aarts, Verhage, Veenvliet, Dolan, & van der Sluis, 2014; Baayen, Davidson, & Bates, 2008; Jaeger, 2008; Lazic, 2010; Locker, Hoffman, & Bovaird, 2007; Payne et al., 2014). The linear mixed-effects model is a special (restricted) case of models that are commonly used in psychophysiology, including repeated measures (mixed-effects) analysis of variance (ANOVA) and ordinary least-squares regression. LMMs are useful for modeling data with complex sampling or clustering, such that observations in the response vector are nonindependent. Statistical dependencies of this sort exist, for example, when observations are drawn in a nonrandom manner (e.g., repeated measurements across trials within the same subjects or spatial correlation across neighboring electrode sites). Such data are common in ERP experiments of sentence processing that often contain multiple sources of statistical dependency (e.g., across words, subjects, channel sites). In these cases, the use of LMMs avoids the need for aggregation across either items (i.e., by-subjects analyses) or subjects (i.e., by-items analyses).

In more general terms, LMMs are useful for (a) modeling an arbitrary number of fixed and random effects (as long as the issue of overfitting is addressed), allowing for the flexible accommodation of many experimental designs and methodologies without having to first transform or aggregate raw data; (b) modeling the influence of predictors at multiple levels of variation (e.g., across subjects, items, time, space) simultaneously; (c) joint modeling of both discrete and continuous variables; (d) modeling data from unbalanced designs with missing data; and (e) explicitly modeling the variance-covariance structure of the data, allowing for violations of sphericity and homogeneity of error variance (see Gelman & Hill, 2007; Snijders & Bosker, 2011). Thus, LMMs offer a tool for fitting models to accommodate arbitrarily complex study designs by modeling of the variance-covariance structure of the data, such that the researcher can abandon the common method of forcing or coercing data to fit within a prespecified model (e.g., ANOVA, regression). Appendix S1 provides supplementary details regarding the linear mixed-effects model, including model fitting, implementation, and interpretation.

One aim of the current study was to demonstrate that linear mixed-effects models are useful tools for testing hypotheses about item-level variation in ERPs and that LMMs can be used to examine continuous item-level dynamics of event-related EEG without the loss of information and precision entailed by averaging and discretizing naturally continuous variables. Toward this aim, our first goal was to replicate the word-position effect (Van Petten & Kutas, 1990, 1991) on the N400 at the level of individual words utilizing item-level modeling of the N400 component (see also Dambacher et al., 2006, for a conceptually related analysis utilizing subject-specific random regression models). Our primary aim in the current study was to then demonstrate that such item-level analyses yield new insights into the sensitivity of the N400 to multiple different sources of natural linguistic variation, by examining the degree to which word-level variation in frequency and orthographic neighborhood size is moderated by accumulating higher-order semantic and syntactic contextual constraints.

Method

Participants

Data were analyzed from 28 participants (13 females; mean age = 20, range = 18–37); 24 of those datasets were previously analyzed in Lee and Federmeier (2009; looking at responses to sentence-final words only). All participants were right-handed monolingual native speakers of English with normal or corrected-to-normal vision. None of the participants had a history of neurological or psychiatric disorders or brain damage. Participants were compensated with course credit.

Materials

Participants read a total of 172 sentences, divided into three conditions: (1) congruent sentences (e.g., *She kept checking the oven because the cake seemed to be taking an awfully long time to bake*), (2) syntactic prose sentences (*She went missing the spring because the court began to be making an awfully poor art to bake*), which provide the same syntactic structure as the coherent items, but with no coherent message-level semantics, and (3) scrambled sentences (*The court the she spring making missing awfully art poor to because an to be went began bake*). Syntactic prose sentences were created by replacing the content words of each congruent sentence with randomly selected words of the same grammatical category from other congruent sentences. Therefore, congruent and syntactic prose sentences were matched in the relative position of closed-class words. Random sentences were created by randomly scrambling the position of the words within each syntactic prose sentence, with the exception of the sentence-final word. Sentences contained, on average, about 14 words ($M = 14.20$, $SD = 3.39$, *range* = 5–27).

Open class words (typically defined as "meaning-bearing" words) included nouns, verbs, adjectives, and derived adverbs (-ly adverbs). Closed class words (semantically sparse words that mainly perform syntactic functions) included words belonging to other lexical classes (e.g., determiners, prepositions, conjunctions, and pronouns). Following the dichotomous assignment of words in Van Petten and Kutas (1991), words of ambiguous class were assigned to the closed-class category. Although identical closed-class words appeared in the three conditions, open-class words were presented across conditions with random selection, but without exhausting all possible permutations, due to limitations in the number of stimuli that could be presented in a single session. Importantly, no differences were found across sentence contexts in any of the target or control variables analyzed in the current study.

Procedure

Participants were seated 100 cm in front of a 21" computer monitor in a dim, quiet testing room. At the start of each trial, a series of plus signs appeared in the center of the screen for 500 ms. After a stimulus onset asynchrony (SOA) ranging between 1,000–1,500 ms (randomly jittered to reduce anticipatory potentials), a sentence was displayed word by word in the center of the screen. Each word was presented for 200 ms, followed by a 300 ms blank screen. To ensure that participants were attending to each word, as well as attempting to integrate each word into a holistic unit, participants

were administered word and sentence recognition tasks. Following each sentence, participants were presented with a probe word and asked to judge whether it had appeared in the preceding sentence. Half of the probe words were new words and half of the probes appeared in the previous sentence. The experimental session was divided into eight blocks. At the end of every two blocks, participants were also administered a brief sentence-recognition test. In total, participants were tested on 96 sentences, half of which were old (drawn in equal numbers from congruent, syntactic prose, and scrambled sentences) and half of which were new (also consisting of equal numbers of each sentence type). New sentences contained some words that the participant actually viewed, making word-level recognition alone insufficient to allow participants to succeed on this test. As the behavioral data have already been reported in Lee and Federmeier (2009), we do not redescribe them here, except to say that the results showed that participants were attending to both individual words and to the sentences as a whole.

EEG Recording and Processing

EEG was recorded from 26 evenly spaced silver-silver chloride electrodes embedded in an Electro-Cap. The sites were midline prefrontal (MiPf), (left and right) medial prefrontal (L/RMPf), lateral prefrontal (L/RMPf), medial frontal (L/RMFr), mediolateral frontal (L/RDFr), lateral frontal (L/RLFr), midline central (MiCe), medial central (L/RMCe), mediolateral central (L/RDCe), midline parietal (MiPa), mediolateral parietal (L/RDPa), lateral temporal (L/RLTe), midline occipital (MiOc), medial occipital (L/RMOC), and lateral occipital (L/RLOc).

All scalp electrodes were referenced online to the left mastoid and re-referenced offline to the average of the right and the left mastoids. In addition, one electrode (referenced to the left mastoid) was placed on the left infraorbital ridge to monitor for vertical eye movements and blinks, and another two electrodes (referenced to one another) were placed on the outer canthus of each eye to monitor for horizontal eye movements. Electrode impedances were kept below 3 kΩ. The continuous EEG was amplified through a bandpass filter of .02–100 Hz and recorded to hard disk at a sampling rate of 250 Hz.

EEG epochs were examined and marked for artifacts (drift, muscle activity, eye blinks, and eye movements). On average, a total of 18% ($SD = 13\%$; range across subjects $= < 1\%$–48%) of words were marked as artifacts and not included in subsequent analyses. We adopted a very conservative approach to removing entire subjects on the basis of artifacts (greater than 50% of data loss). This approach resulted in removing four subjects from analyses. Analyses were conducted via maximum likelihood estimation on all available data (Graham, 2009; Little & Rubin, 2002).[1] This method accommodates unbalanced designs that arise from artifact rejection by using all available data to estimate parameters, such that highly influential individual random effects (e.g., subjects or words) with fewer observations are shrunk toward the population average (see Appendix S2).

To examine word-by-word variation in N400 activity, measurements of mean N400 amplitude were collected at the level of individual words from the raw EEG. Epoched (from 100 ms prestimulus to 920 ms poststimulus) unaveraged EEG for each word was treated as the critical data. Item-level amplitudes were measured from all words in each sentence, including the sentence initial and final words. Some previous studies have removed these items before calculating grand averages (e.g., Van Petten & Kutas, 1991). Nevertheless, we found that removal of these items did not alter the pattern of results. Thus, all possible words were included in the analyses. Because our sentences varied substantially in length, sentence-final words were distributed across different word positions from sentence to sentence. Following baseline correction (conducted by subtracting the 100 ms prestimulus baseline period), measurements of mean amplitude were taken within a predefined N400 epoch (300–500 ms) at each channel separately for each word, after applying a digital low-pass filter of 30 Hz. The resulting dataset includes measurements of mean amplitude within the N400 latency band separately for each word, channel, and participant.

Data Analysis

Analyses of word-level N400 amplitudes were conducted using linear mixed-effects models via restricted maximum likelihood estimation. All analyses were conducted with the *lme4* package (Bates, Maechler, Bolker, & Walker, 2014) in the *R* language for statistical computing (R Core Team, 2014). Appendix S1 contains supplementary information for fitting the models used in the current article.

We defined the random-effects structure of our models to represent the inherent experimental design and nested sampling structure of our data (cf. Barr, Levy, Scheepers, & Tily, 2013). Thus, variance across subjects, items, and channels were modeled as random intercept terms in the statistical model. Preliminary models also included a random intercept for sentence. However, this parameter was estimated at zero, indicating that there was little unique residual variation in N400 amplitude across sentences after accounting for word-level variability. Analyses of N400 effects were conducted across eight a priori chosen centro-parietal electrode sites (LMCe, RMCe, LDCe, RDCe, LDPa, RDPa, MiCe, MiPa), where N400 effects are typically largest (with the exception of the distributional analyses, where models were fit across all scalp channels). In the analyses of contextual and lexical influences, only open-class words were considered, as effects of frequency and sentential context on the N400 are largest within these words (Van Petten & Kutas, 1990, 1991).

Predictors of word-level variance included sentence context (SC: congruent, syntactic prose, or random), word position (WP), word frequency, and orthographic neighborhood. Word frequency (log transformed) was derived from the Hyperspace Analog to Language (HAL) norms from the English Lexicon Project, and orthographic neighborhood size was derived from the orthographic Levenshtein distance 20 (OLD20) measure (Yarkoni et al., 2008) from the English Lexicon Project (see Balota et al., 2007). OLD20 reflects the mean distance (in number of steps) from each word to the 20 closest Levenshtein neighbors in the lexicon. Levenshtein distance (Levenshtein, 1966) is the minimum number of substitutions, insertions, or deletion operations required to turn one word into another. Thus, words with higher OLD20 scores are considered orthographically sparse (have relatively fewer neighbors), whereas words with lower OLD20 scores are considered orthographically dense (have relatively more neighbors). This measure is, thus, negatively correlated with traditional measures of

1. In contrast to single and multiple imputation-based methods (which aim to fill in missing data), or methods that result in complete-case data through data deletion methods (which result in biased estimates), ML methods allow for the modeling of incomplete/unbalanced data by finding parameters that maximize the likelihood (in an iterative manner; see Appendix S1) using all available data for those parameters. Notably, under the assumption that data are missing at random (conditional on model parameters) or missing completely at random, ML estimates are not biased by data missingness (see Graham, 2009; Molenbergs & Kenward, 2008; Schafer & Yucel, 2002).

Table 1. *Lexical Characteristics of Open-Class Words in Model 2*

			Correlations			
	Mean (*SD*)	Range (Min–Max)	1	2	3	4
1. Word frequency (log)	11.11 (2.16)	4.06–15.68				
2. Orthographic N (OLD 20)	1.77 (0.62)	1.00–5.50	–.45			
3. Word length (Characters)	5.12 (1.91)	2–14	–.55	.73		
4. Concreteness	3.07 (1.15)	1.12–5.00	–.48	.01	.10	
5. Word position	7.46 (4.85)	1–27	–.08	–.02	–.02	.01

neighborhood size, such as Coltheart's *N* (e.g., the number of words that can be obtained by changing one letter while preserving the identity and positions of the other letters; Coltheart et al., 1977). Word length, sentence length, and concreteness ratings were also used as control variables in some analyses (see below). Concreteness ratings were drawn from a recent large-scale norming study (Brysbaert, Warriner, & Kuperman, 2014). Table 1 presents descriptive information and correlations among these lexical variables.

Ordinal word position has a skewed distribution because it is a cumulative measure (i.e., all sentences have at least 5 words, but few sentences have more than 20 words). Several data transformations (e.g., natural log transformation, Box-Cox power transformation, sentence-mean centering) were conducted on word position, but all analyses resulted in the same pattern of results (cf. Kuperman, Dambacher, Nuthmann, & Kliegl, 2010). Thus, for transparency of interpretation, word position was coded as the ordinal position from the beginning of each sentence.

Because sentence context has three levels, the congruent condition was treated as a reference group to form two contrasts: Contrast 1 (SC1): syntactic prose vs. congruent; Contrast 2 (SC2): random vs. congruent. These contrasts were used in all models unless otherwise noted. All other variables were treated as continuous effects. All continuous variables were standardized before analysis in order to center and scale the predictors, which reduces unessential multicollinearity and simplifies interpretation of parameter estimates in the presence of higher-order interactions. The interpretation of fixed-effect parameter estimates is analogous to the interpretation of regression weights in the linear regression model. Thus, note that important concepts necessary for interpreting higher-order interactions in linear regression models (e.g., effects of centering, contrast coding, the principle of marginality, interactions between dichotomous and continuous covariates; Cohen, Cohen, West, & Aiken, 2003; Hayes, 2013) also hold for the fixed-effects in linear mixed-effects models.

For continuous variables, parameter estimates reflect change in mean amplitude per standard deviation change in the variable. For dichotomous variables, effect sizes reflect change in mean amplitude between the reference and contrast group. Parameter estimates for higher-order interactions (including continuous and dichotomous variables) reflect the magnitude of the effect of one of the independent variables on a dependent variable as a function of two (or more) independent variables (interpreted as moderator variables). Conditional plots probing key higher-order interactions are included to aid in interpretation (cf. Bauer & Curran, 2005; Curran, Bauer, & Willoughby, 2004). In addition, when higher-order interactions were reliable, they were further probed by fitting separate models as a function of sentence context.

Further specification of the random effects structure was modeled following the recommendations of Barr and colleagues (Barr, 2013; Barr et al., 2013). Initial models were fit with random slope parameters for all corresponding within-subject effects warranted by the design (i.e., a fully maximal random-effects structure). Note that because our word-level effects of interest were not experimentally crossed, but rather properties of the words (e.g., position in the sentence, frequency), by-word random slopes for word-level predictors were not considered. Because of the massive number of variance-covariance parameters to be estimated, the maximal models unsurprisingly failed to converge to a proper solution. Simplified random effects structures were fit, aiming to reduce overfitting of the random effects structure (cf. Bates, Kliegl, Vasishth, & Baayen, 2015). In simplified models, by-channel random slope parameters were estimated at zero, resulting in failures to converge to an optimal solution. This likely reflects the limited variance in effects across the selected centro-parietal channels due to volume conduction. Therefore, random slopes of effects across channels were not fit in final models. The final converging models included by-subjects random-slope variance estimates for the critical highest-order interactions in the models, which results in a balanced trade-off between model variance-covariance matrix overfitting and deriving *SE* estimates that are appropriately conservative (see Barr, 2013; Barr et al., 2013).

First, we present a model testing the degree to which N400 amplitudes vary with word position as a function of sentence context (Model 1). The aim of this analysis is to replicate the word-position context effect, to show that effects of sentence context on the N400 can be detected at the level of individual words. Following this, we simultaneously examine the impact of word frequency and orthographic neighborhood on N400 activity and, in particular, assess how the impact of these lexical variables is moderated by sentence context and word position (Model 2). Effects sizes are presented as model-derived fixed-effect parameter estimates (i.e., regression weights), along with corresponding 95% profile likelihood confidence intervals for statistical inference (Cumming, 2014). Parameters with confidence intervals that do not contain zero can be interpreted as statistically significant following traditional null-hypothesis significance testing. Comparative and absolute fit statistics from these models are presented in Table 2 (see Appendix S3 for more details). These include the likelihood ratio test, Akakie Information Criteria (AIC), Bayesian Information Criteria (BIC), and two approximate R^2 measures—pseudo R^2 and conditional R^2 (Johnson, 2014; Nakagawa & Schielzeth, 2013; Singer & Willet, 2003). A "null" model, which includes only an intercept parameter plus the random-intercept structure, is presented in Table 2 as a point of comparison for fit statistics. See Appendix S3 for further information on fit indices.

Results

Incremental Effects of Semantic and Syntactic Context

First, a model was fit testing the effects of word position (WP) and sentence context (SC) on N400 amplitudes to all open-class words

Table 2. *Fit Statistics From Models 1 and 2*

	−2LL	$\chi^2(df)$	*p*	AIC	BIC	pR^2	cR^2
Null Model	1,450,641			1,450,651	1,450,703	5.71%	7.55%
Model 1	1,449,738	903 (7)	<.001	1,449,762	1,449,886	6.17%	7.89%
Model 2	1,448,750	1,891.7 (25)	<.001	1,448,810	1,449,118	6.69%	8.23%

Note. The Null Model is a nested model containing only a fixed-intercept term plus the random intercept terms, used to assess baseline fit. Models 1 and 2 are defined on page 5. −2LL = −2 times the log of the likelihood for the model; χ^2 = deviance statistic between Model 1/2 and the Null Model; AIC = Akakie Information Criteria; BIC = Bayesian Information Criteria; pR^2 = pseudo R^2 (Singer & Willet, 2003); cR^2 = conditional R^2 (Johnson, 2014; Nakagawa & Schielzeth, 2013). See Appendix S3 for further information on fit indices.

(Model 1). Of key interest here is the test of whether word position interacts with each sentence context contrast. We found reliable WP x SC1 (b = –.46; 95% CI = [–.645, –.27]) and WP x SC2 interactions (b = –.50; 95% CI = [–.65, –.35]), indicating that the word-position effect was reliably larger in congruent sentences than in syntactic prose or random sentences. Figure 1a plots the best-fit linear regression lines for word position in each sentence context. As can be seen in Figure 1, there was a robust linear relationship between word position and N400 amplitude in congruent sentences, such that increasing word position was associated with reduced (more positive) N400 amplitudes, an effect that was not present in syntactic prose or randomly shuffled word strings. Figure 1b plots grand-average ERPs, illustrating the reliable word-position effect on N400 amplitude in congruent sentences. Two-word bins are plotted across six electrode sites, color-coded by ordinal word position within sentences. Although there was some slow variation across the entire waveform from word to word, large systematic variation as a function of word position was seen only within the N400 epoch.

Parametric (via modeling higher-order polynomial trends) and nonparametric (via generalized additive models) nonlinear trends of word position were also considered in separate analyses. While there was some indication of nonlinearity across word position, the overwhelming trend was for a reduction in N400 with increasing word position that was well captured by a simple linear trend (c.f., Marslen-Wilson & Tyler, 1980). Thus, linear trends were used in the current study, striking a balance between fit to the data and parsimonious inferential interpretation of model results, especially when considering higher-order interactions in later models.

Figure 1c presents the scalp distribution of the word-position effect in congruent sentences. We computed channel-specific predictions of the linear effect of word position on N400 amplitudes by estimating the best linear unbiased predictors (*BLUP*s) (see Appendix S2) of this effect on N400 amplitude across all scalp channels in a separate model fit to all open-class words within congruent sentences only. These estimated effect sizes represent the per-word decrease in N400 amplitude across word position, separately for each scalp channel. More negative values indicate a larger estimated decrease in N400 amplitude per unit increase in word position. A scalp topography map of the word-position effect was created by mapping the channel-specific data to a two-dimensional circular head and using spherical spline interpolation of values between

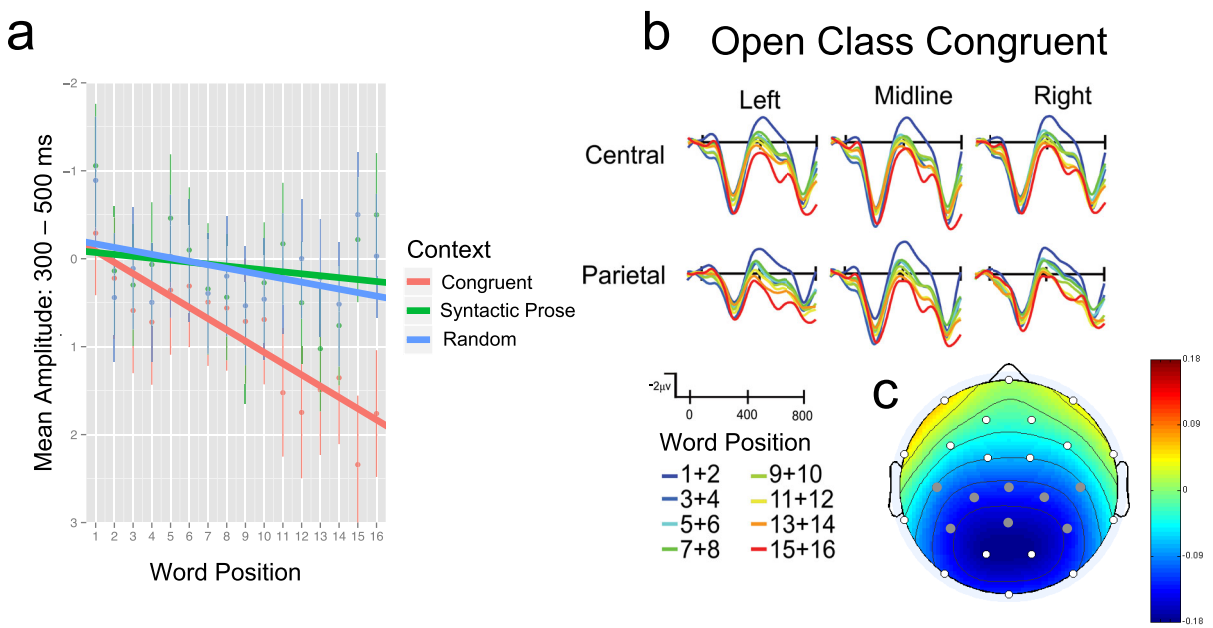

Figure 1. a: Linear word-position effects on single-word ERPs to open-class words in the N400 epoch (300–500 ms) plotted separately for each sentence context. Error bars reflect the between-subject standard error of the mean computed across all subjects, words, and channels at each word position. b: Grand-average ERPs illustrating word-position effects for open-class words in congruent sentences. Two-word bins are presented, color coded by word position, over six central parietal electrodes. Negative is plotted up. c: Scalp topography of the best linear unbiased estimates of word-position effects on N400 amplitude for open-class words in congruent sentences (see text for details). Electrode channel sites with gray circles are those for which data was included in the single-word mixed-effect models.

channels on a fine Cartesian grid via the *topoplot* function in the EEGLAB (Delorme & Makeig, 2004) toolbox for MATLAB (The MathWorks, 2014). The resulting figure represents the spatial distribution of the effect size (linear slope estimate) of word position on N400 amplitudes across the scalp. Note that this is similar to plotting the distribution of a difference wave to visualize the scalp topography of a particular experimental effect. In this case, because the variable of interest is continuous and approximately linear, the correct corresponding plot of an effect would be the linear slope estimate, as shown here. This figure clearly shows that the word-position effect follows a characteristic N400 scalp distribution, with the largest effects over centro-parietal electrode sites.

The analyses reported above illustrate that semantic modulation of N400 activity can be detected at the item level. Indeed, Figure 1b clearly shows word-position effects on the grand-averaged N400, consistent with the findings from the item-level models. However, this figure also shows that there is not complete equivalence across word positions in the baseline. To determine whether the word-position effects reported above are driven by confounding factors early in the waveform that may be influencing component measurement at the item level (e.g., slow potentials, early sensory/perceptual differences, preceding component overlap, or preparatory activity), a control model was tested on an early period in the event-related item EEG. A model was fit that was identical to the initial WP x SC model, except that it was fit to word-level mean amplitudes in the period from 0 ms to 200 ms poststimulus onset. This period is the same size as the N400 latency measurement window (300–500 ms), but is one in which semantic influences would be unexpected (Laszlo & Federmeier, 2014). We found no reliable interactions nor any evidence for reliable word-position effects within any sentence type, suggesting that the N400 word-position effect is not driven by early or baseline item-to-item fluctuations in amplitude.

Lexico-Semantic Modulation of the N400 in Sentence Contexts

In our initial analysis, we showed that we could replicate the word-position context effect with N400 amplitudes measured at the level of individual words, thus illustrating the validity of an item-level modeling approach. Our next aim was to examine the degree to which frequency and neighborhood effects simultaneously contribute to N400 amplitude during sentence reading and to study how semantic and syntactic contexts influence lexico-semantic processing as indexed by the N400. Notably, such analyses are not possible via traditional aggregate approaches that average across multiple items and discretize continuous variables.

A model was fit to the data with orthographic neighborhood, word frequency, word position, sentence context, and their interactions as predictors of N400 amplitude (Model 2). The aim of this model was to test the effects of accumulating sentence contexts on lexical processing, as indexed by frequency and neighborhood effects on the N400. Thus, we were testing for the presence of three-way interactions between sentence context, word position, and frequency/neighborhood. Effects are adjusted for variation in semantic concreteness, word length, and sentence length. Concreteness and length are included as control variables because they are correlated with both frequency and orthographic neighborhood. Moreover, concreteness has been shown to be a strong independent predictor of N400 amplitude in a recent single-item ERP investigation of word recognition (Van Petten, 2014). Sentence length is also included as a covariate, because variability in overall length may contribute to the strength of word position as a moderator of lexical effects.

Figure 2 presents the fixed-effects parameter estimates and corresponding 95% confidence intervals from the linear mixed-effects model corresponding to this analysis. Of critical interest in this model is the degree to which frequency and orthographic neighborhood effects are moderated by sentence context and word position (i.e., three-way interactions with sentence context and word position). As seen in Figure 2, there were reliable three-way interactions between word position, frequency, and both sentence context contrasts. This interaction is presented graphically in Figure 3a, which depicts the partial-effects plot (see Preacher, Curran, & Bauer, 2006) of word frequency on N400 amplitude at conditional levels of word position (25th, 50th, and 75th percentiles) for congruent, syntactic prose, and random sentences.

To further probe this three-way interaction, individual models were fit testing the Frequency x Position interactions in each sentence type. For congruent sentences, there was a robust effect of word frequency at the beginning of the sentence, which was reduced as word position increased, yielding a reliable WP x Frequency interaction (b = –.57; 95% CI = [–.33, –.81]). This interaction was not statistically significant in syntactic prose sentences (b = –.10; 95% CI = [–.36, .16]) or in random sentences (b = –.16; 95% CI = [–.36, .04]). Collectively, these findings suggest that accumulating message-level semantic constraints reduce the influence of word frequency on semantic processing.

There was no evidence for three-way interactions between orthographic neighborhood size, context, and word position (see Figure 2).[2] Figure 3b presents the partial-effects plot (see Preacher et al., 2006) of orthographic neighborhood on N400 amplitude at conditional levels of word position (25th, 50th, and 75th percentiles) for congruent, syntactic prose, and random sentences. In fact, there was no reliable two-way interaction between word position and orthographic neighborhood in any of the three sentence contexts, suggesting that neighborhood effects are invariant in magnitude across word positions within a sentence. There was, however, evidence for a reliable, but small, two-way interactions between orthographic neighborhood and sentence context in the overall model (Figure 2), such that neighborhood effects were slightly larger in congruent sentences (b_C = .44, 95% CI: [.07, .62]) than in syntactic prose (b_J = .28, 95% CI: [.02, .54]) or random sentences (b_R = .32; 95% CI: [.06, .58]). However, neighborhood effects reliably predicted N400 amplitude across all sentence context types. Thus, in contrast to the effects of frequency, it appears that orthographic neighborhood remains a reliable predictor of N400 amplitudes in the face of increasing message-level semantic constraints.[3]

2. The correlation between word length and orthographic neighborhood is driven by the fact that words that are quite long tend to have a sparse orthographic neighborhood space. Given the high degree of correlation between word length and ON, there is concern about collinearity influencing model parameters. Therefore, we conducted a follow-up analysis aimed at examining the effects of ON in a model without longer words (that necessarily contain fewer neighbors). This analysis was conducted on a restricted dataset excluding words longer than eight characters. Importantly, the pattern of results remains the same: We still find reliable relationships between ON and N400 amplitude in the restricted dataset. Indeed, the effect of orthographic neighborhood was larger overall in the model removing longer words than in our full model.

3. Models were also fit using Coltheart's N as our measure of orthographic neighborhood. This model produced the same pattern of findings (with ON showing robust effects across all word positions in each sentence type and no interactions with sentence context or word position).

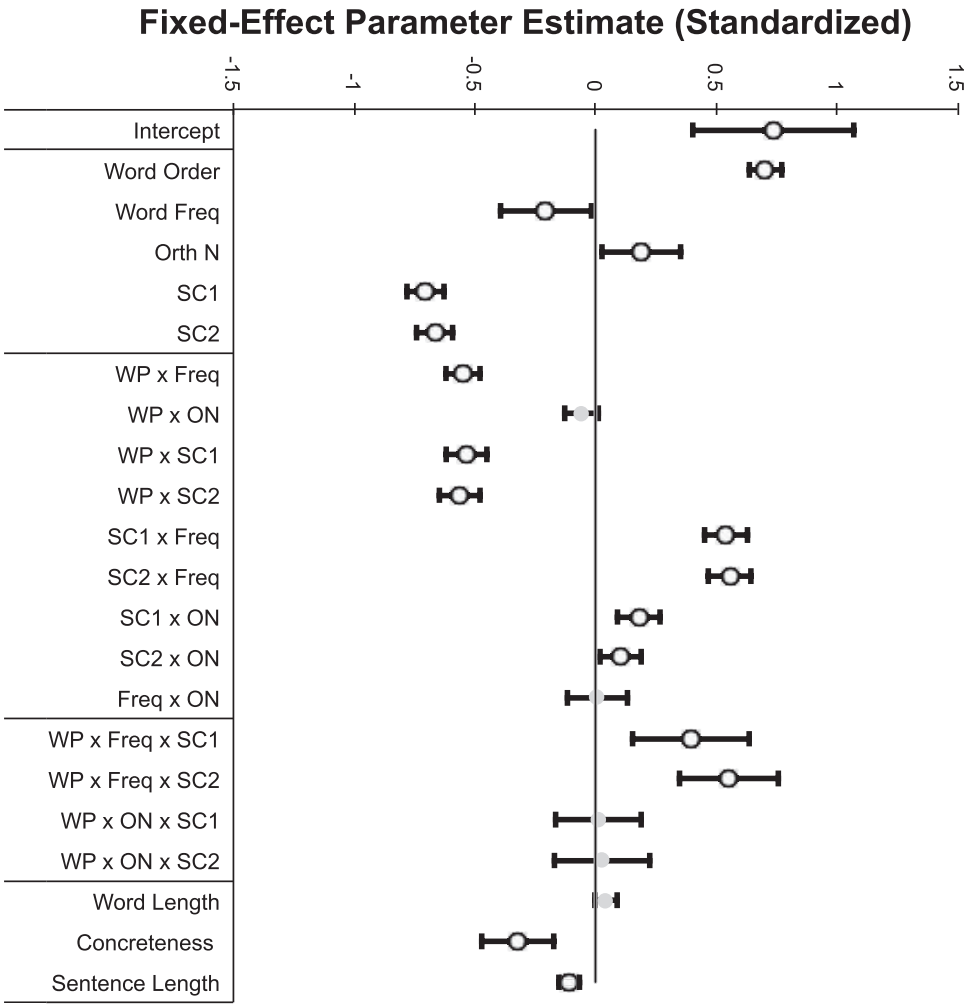

Figure 2. Fixed-effect parameter estimates and corresponding 95% profile confidence intervals from Model 2. *Note*: Estimates with intervals containing 0 (gray circles) do not meet traditional levels of statistical significance.

Effects of Semantic and Syntactic Context on Closed-Class Words

Van Petten and Kutas (1991) found a main effect of sentence context on N400 amplitudes to closed-class words, such that amplitudes were reduced in syntactic prose sentences relative to random sentences, but this effect did not interact with word position. They argued that syntactic context exerted local constraints on closed-class words, which were restricted to minor syntactic constituents, but that this did not increase in strength over the course of the sentence. To test the degree to which accumulating syntactic and semantic context modulated N400 amplitudes to closed-class words, we conducted a follow-up analysis testing the effects of word position and sentence context on N400 amplitudes to all closed-class words, using the same structure as Model 1. Interestingly, we found a reliable WP x SC1 interaction ($b = –.42$; 95% CI = [–.62, –.23]), indicating that closed-class words in syntactic prose showed a larger reduction in amplitude as a function of word position than in the congruent sentences.

Figure 4a plots the best-fit linear regression lines for word position in each sentence context. As can be seen, increasing word position was associated with differentially reduced (more positive) N400 amplitudes to closed-class words in syntactic prose sentences only. Figure 4b plots grand-average ERPs illustrating the word-position effects on N400 amplitudes to closed-class words in syntactic prose. As can be seen, the N400 becomes more positive with increasing word position, although this effect is smaller than the effects of semantic context on open-class words. Figure 4c presents the scalp distribution of the word-position effect on closed-class words in syntactic prose sentences. The word-position effect follows a typical N400 scalp distribution, with a slight right lateralization (cf. Kutas & Hillyard, 1982).

Discussion

The goal of this study was to probe the nature and scope of the incremental effects of semantic and syntactic context on lexical processing during sentence comprehension. Our approach involved the application of an emerging analytical technique, which has not been widely applied to ERP data, in order to uncover word-level N400 dynamics via item-level measurement and analysis of N400 amplitudes, without aggregating EEG across individual subjects or individual items. Our findings provided a replication of the general effects of incremental sentential context on the N400, as well as novel extensions of our understanding of how the buildup of semantic and syntactic constraints impacts word processing.

In particular, we replicated the finding that N400 amplitudes to open-class words are reduced with ordinal word position in congruent sentences only, as reported by Van Petten and Kutas (1991). In

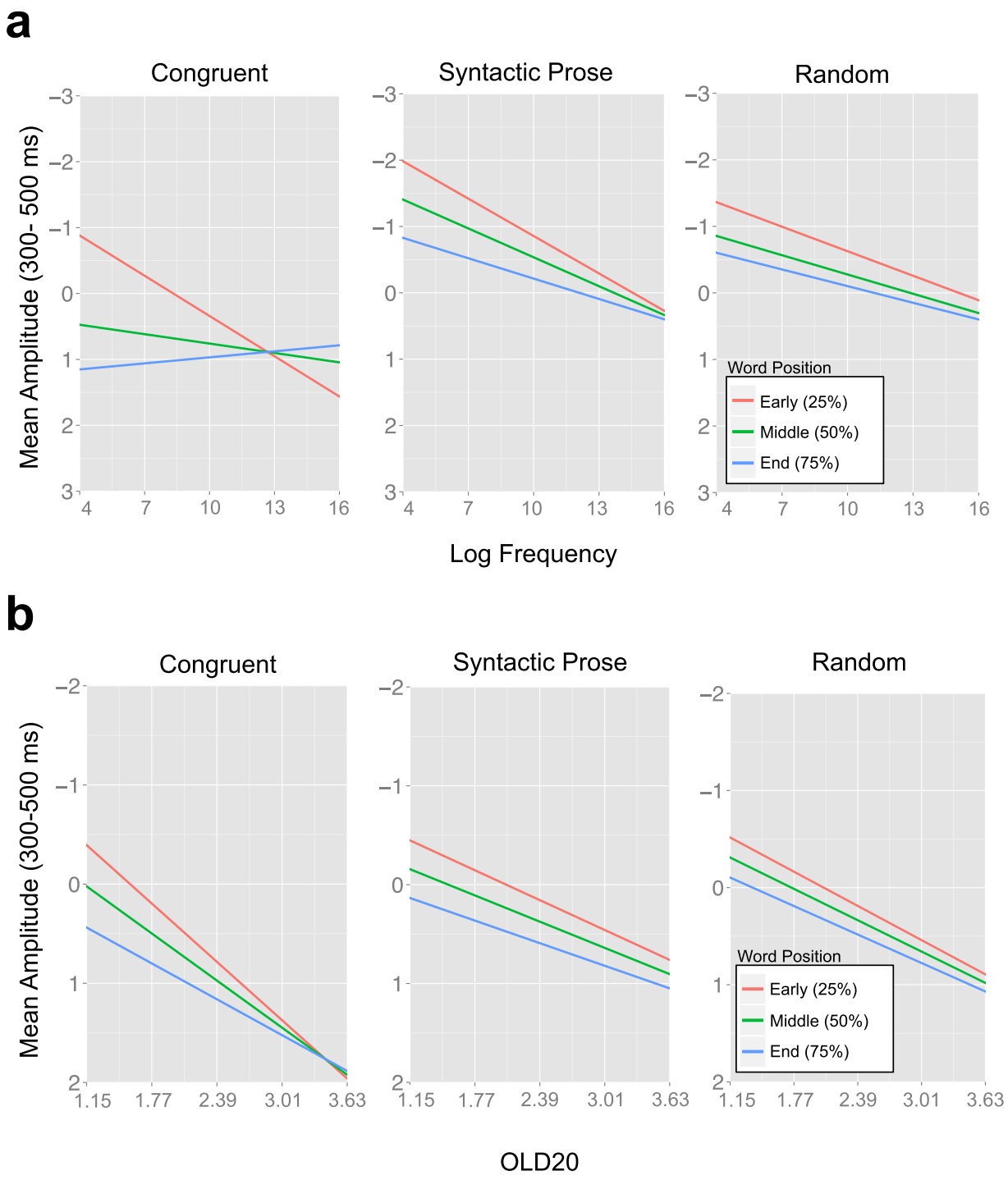

Figure 3. a: Model estimated partial-effects plots of the Frequency × Word position × Context interaction. b: Model estimated partial-effects plots of the Orthographic neighborhood × Word position × Context interaction.

syntactic prose and sentences with randomly shuffled words, there was no evidence that word position modulated N400 amplitudes to open-class words, implicating the accrual of sentence-level semantic context as responsible for the observed word-position effects. By replicating this classic effect utilizing an item-level analysis, where no averaged ERP components were computed or directly measured, we clearly show that functional changes in underlying ERP components can be reliably detected in a single statistical analysis from the event-related EEG without signal averaging across items, at least in the case of N400 amplitudes, which have a uniform stimulus-locked temporal signature. Further evidence for the efficacy of this approach in detecting N400 effects comes from the scalp distributions of the item-level word-position effects, which show a clear centro-parietal distribution, consistent with the observed scalp distribution of the N400 in averaged ERPs (see Figures 1c and 4c).

Given that we could replicate this effect, our principal aim in the current study was then to examine the degree to which the

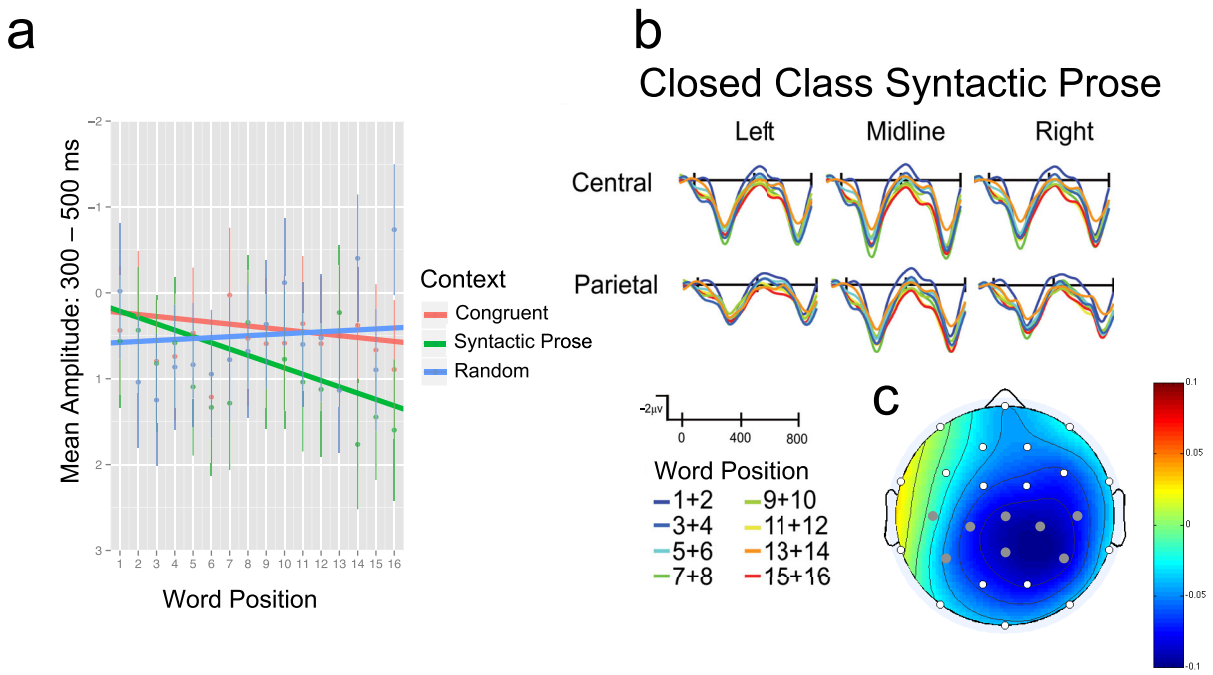

Figure 4. a: Linear word-position effects on single-word ERPs to closed-class words in the N400 epoch (300–500 ms) plotted separately for each sentence context. Error bars reflect the between-subject standard error of the mean computed across all subjects, words, and channels. b: Grand-average ERPs illustrating word-position effects for closed-class words in syntactic prose. Two-word bins are presented, color-coded by word position, over six central parietal electrodes. Negative is plotted up. c: Scalp topography of the best linear unbiased estimates of word-position effects on N400 amplitude for closed-class words in syntactic prose (see text for details). Electrode channel sites with gray circles are those for which data was included in the single-word mixed-effect models.

contributions of frequency and orthographic neighborhood to word processing are modulated by the availability of semantic and syntactic constraints. Therefore, we examined how the incremental buildup of context interacted with lexical processing by simultaneously modeling the effects of orthographic neighborhood and word frequency on the N400 to individual words, across sentence types that differed in the availability of those constraints and over word position within a sentence as those constraints built up. The results from these analyses highlight that frequency and orthographic neighborhood effects on the N400 do not respond similarly to accumulating sentential context.

At the beginning of congruent sentences, clear frequency effects emerged, such that the N400 was larger to less frequent words. However, with accumulating semantic context only, effects of frequency were reduced in magnitude, replicating the findings from Van Petten and Kutas (1991). In contrast, orthographic neighborhood showed a different pattern of sensitivity to sentence context. First, we replicated previous findings showing an association between neighborhood size and N400 amplitude, with larger N400s to words with many neighbors compared to those with fewer neighbors (Holcomb et al., 2002, Laszlo & Federmeier 2007, 2008, 2009). Our findings extend earlier work on neighborhood effects by showing that (a) orthographic neighborhood effects can be observed continuously at the level of individual words in sentences, in the absence of signal averaging across subjects or items, (b) orthographic neighborhood effects are observed invariant of word position, and (c) accumulating semantic context does not eliminate neighborhood effects, unlike effects of frequency.

Most notably, this study is the first to directly compare the effects of sentence context on frequency and neighborhood effects simultaneously across varying context types. Federmeier and Laszlo (2009) argued that frequency and orthographic neighborhood effects reflect, respectively, the dynamics of information use and information structure during visual word recognition. Our findings were consistent with this model. We found that supportive sentence contexts reduced the impact of frequency over the course of a sentence, whereas orthographic neighborhood effects persisted in congruent sentences. Indeed, semantically supportive sentence contexts appeared to slightly increase the magnitude of the neighborhood effect on the N400, with a small but reliably larger neighborhood effect in congruent sentences relative to syntactic prose. In syntactic prose sentences, neighborhood effects were slightly smaller on average—but still present—suggesting that this atypical syntactic structure may have interfered to some degree with early aspects of lexical processing (i.e., initial spreading activation to orthographically similar word form representations). This finding is additionally consistent with the claim that in syntactic prose items, readers may be strategically shifting attention away from open-class words, instead focusing more strongly on the overall syntactic structure (see discussion below).

Whereas orthographic neighborhood reflects the degree to which visual word representations are structured by orthographic similarity, frequency effects appear to represent transient "baseline" activation states of semantic memory based on prior experience. Such activation states are likely to be malleable—as supported by findings that even for out-of-context words, frequency effects on the N400 (and on behavioral indices of word processing) vary with task demands (see discussion in Fisher-Baum et al., 2014). Similarly, we would expect that the buildup of message-level constraints would adjust these activation states away from their baselines, thus reducing the influence of frequency on the N400.

We also found, in contrast to Van Petten and Kutas (1991), that N400s to closed-class words were reduced with accumulating syntactic context. In their study, there was an overall main effect of sentence context such that syntactic prose sentences showed reduced N400 effects overall compared to congruent words. Van Petten and Kutas argued that syntactic context did exert constraints on closed-class words, but that these were locally bounded within phrases or clauses (e.g., a preposition predicting a subsequent determiner as in "*He was in the house"*). Such local constraints would not be expected to increase with accumulating syntactic context.

It is possible that such syntactic prediction is task specific or strategic in nature, guided in part by the constraints provided by the sentence context. In our sentences, which included (on average) longer sentences than Van Petten and Kutas (1991), it is not immediately obvious that syntactic prose sentences are incoherent, sometimes taking several words for a reader to realize that a sentence provides only syntactic cues. Thus, the word-position effect on closed-class words may reflect the shifting of attention explicitly toward syntactic structure. This is consistent with findings from an event-related fMRI study by Friederici, Meyer, and von Cramon (2000), who found evidence that relative to normal prose, so-called "Jabberwocky" sentences differentially increased activation in anterior and posterior temporal regions implicated in syntactic processing (see also Mazoyer et al., 1993). Because syntactic prose lacks semantic cues, the syntactic system may be engaged to a greater extent, for example to monitor incoming word-order information during syntactic structure building over the course of the sentence.

Thus, these effects in the syntactic prose condition may reflect task-specific changes in processing (cf. Kaan & Swaab, 2002). To the extent that syntactic prose acts as a task set (which has been shown to modulate the N400; Fischer-Baum et al., 2014), readers may become more biased toward the overarching syntactic structure as syntactic prose sentences unfold in time. When only syntactic information is available, the system may shift focus strategically to those features of the sentence that are consistent and predictable within the limits of this context. Importantly, these findings are consistent with a growing view of the N400 as part of a highly interactive system that immediately takes advantage of all available information in parallel to guide word processing (Kutas & Federmeier, 2011; Laszlo & Federmeier, 2008).

There are a number of reasons that our study may have resulted in a more "global" effect of syntactic context, reflecting shifts in attention to syntactic information with accumulating syntactic context. One possibility is that the behavioral task used in the current study biased attention away from word processing and more toward structural processing. As part of our offline comprehension assessment, we included a delayed sentence-recognition task in addition to a similar word–recognition task as used by Van Petten and Kutas. It is possible that the sentence-recognition task encouraged participants to allocate more attention to the global sentence structure, especially in the case wherein only syntactic context was available. In addition, our sentences included a much larger range of sentence lengths, with some sentences spanning more than 20 words. It is possible, then, that the increased context afforded by these sentences allowed for an enhanced appreciation of the global syntactic structure, leading to increased featural preactivation for closed-class words. Indeed, our longer sentences may have afforded possibilities for global syntactic predictions that spanned multiple words and syntactic boundaries (e.g., *She was so scared after the football that she managed to weep*).

An alternative explanation for the difference between our findings and those of Van Petten and Kutas (1991) centers around our analytical methodology. By analyzing the unaggregated event-related EEG, and modeling word-position effects continuously, it could also be that we had greater power to detect these subtler effects in syntactic prose. Discretizing the word-position effect (i.e., reducing the levels of a variable by aggregating adjoining values/levels) in order to compute by-subject ERPs in the original study may have distorted these overall small but reliable effects that were revealed in our study that utilized an analytical method that more closely matched the underlying structure of the data. However, this appeal to power does not explain the lack of difference between congruent and random sentences in the current study.

Linear mixed-effects models and related item-level methods (e.g., generalized additive mixed models, Tremblay & Newman, 2015; rERPs, Smith & Kutas, 2015a, 2015b; ERP-images, Delorme et al., 2014) offer a number of advantages in the analysis of ERP and EEG data. We have argued that LMMs are useful for modeling item-level dynamics that would be lost in traditional averaging, modeling of data with unbalanced observations across experimental units, examining continuous covariates at multiple sources of variation in the data (e.g., interactions between trial-level variables and individual difference factors), and the treatment of scalp distributional effects.

However, there are limitations to this approach and areas where future work is necessary. First, an exploratory analysis of the full ERP waveform (and the effects of covariates on the entire ERP waveform) is not easily handled within the LMM framework used in the current study. We aimed to model the scalp-recorded N400, which has very well characterized spatial (centro-parietally maximal) and temporal (peak amplitudes between 300–500 ms) features. However, the choice of time-window is complicated in exploratory analyses, where a clear component may not be selected a priori. This may be mitigated by conducting initial exploratory analyses via traditional ERPs, or by using visualization methods capable of probing single-trial ERP dynamics, such as the use of ERP images (Delorme et al., 2014) or through two-stage regression-based estimation of ERP waveforms (rERPs), as described by Smith and Kutas (2015a,b). These graphical methods could be combined with linear mixed-effects models for statistical inference in order to aid in fully characterizing effects of continuous covariates on ERPs (but see Luck [2014] for issues surrounding bias in visualizing results to select analysis parameters).

We also focused on mean amplitude measures in the current report. Because the expected value is a linear operator, there is no concern in inferential issues surrounding measuring mean amplitude from single trials. However, latency-based measures do not share this same property, as there is not equivalence between the mean of item-level latency estimates and the latency of a mean ERP component. In particular, noise in the trial-level EEG data precludes the simple measurement of trial-level EEG. Future work would benefit from developing methods to couple single-trial denoising methods (Ahmadi & Quiroga, 2013; Mouraux & Iannetti, 2008; Quiroga & Garcia, 2003) with linear-mixed effects models to study single-trial variation in latency-based measures. Nevertheless, we believe that the linear mixed-model framework offers a number of useful tools for modeling event-related EEG and ERP data within a familiar analytical framework that encompasses all of the tools of traditional ERP analysis, while expanding the toolbox to allow for more flexible and appropriate analysis of electrophysiological data.

Last, we did not find that our models explained a substantial amount of overall variance in the single-item EEG (see Table 2). The goal of the current article was to test a set of a priori confirmatory hypotheses, rather than to fit a statistical model that best explained all the variation in the single-trial EEG. Indeed, modeling item-level EEG relative to ERPs will invariably result in substantial differences in the perceived explained variance. By modeling the raw EEG, we necessarily decrease the signal-to-noise ratio relative to ERPs, as variability in the raw EEG across the single-item amplitudes is an order of magnitude larger than what is seen in ERPs. However, signal averaging does not just eliminate pure noise sources, but it also eliminates interesting item-to-item component variability (as we have demonstrated in the current article). The trouble with comparing the absolute-fit performance of models that do a first stage of data reduction (e.g., through averaging) is that information about the uncertainty around those averages is not carried over from the first stage (e.g., average across subjects) to the next (e.g., data analysis), thus inflating estimates of fit. However, LMMs maintain all sources of variability in the model, across item, subject, and channel. Thus, absolute-fit stats will be more modest, but they are also a better reflection of reality.

Indeed, other methods have been used to examine item-to-item (word-to-word) variation in N400 amplitude other than those used here (Laszlo and Federmeier, 2014; Van Petten, 2014). These studies averaged data across subjects separately for individual items, creating word-level averages (similar to an F2-ANOVA common in behavioral psycholinguistics). While this method is useful in some designs (and is likely to yield very similar results to the method utilized here), we believe that analysis of the unaggregated data in a multilevel model is more flexible and offers a number of benefits relative to the F2 approach. Specifically, the mixed-effects model allows for (a) explicit treatment of data in unbalanced designs; (b) the capability of modeling variables at multiple levels in a hierarchy (sentence, word, subject), as well as interactions between these levels; (c) computing a more accurate representation of the degree of variability of single-trial data, resulting in appropriately conservative test statistics; and (d) the capability of modeling single-item effects in partially crossed designs where there may not be a simple mapping across single items for creating across-subject averages, as may be quite common in corpus analyses.

It is worth noting that the methods used in the current study (single-trial measurement and utilization of random effects to control for statistical dependency across subjects, items, and channels) could be combined with other nonparametric and nonlinear approaches (e.g., single-trial EEG classification: Blankertz, Lemm, Treder, Haufe, & Müller, 2011; generalized additive models: Tremblay & Newton, 2015) to build predictive models, which would be of great use in other research settings (e.g., brain–computer interfacing). Thus, LMMs can be used to aid in improved statistical inference in ERP/EEG research, as well as in machine-learning and predictive modeling applications (see Breiman, 2001; Shmueli, 2010).

In conclusion, the current study provides a demonstration of the usefulness of item-level analyses for investigating the multiple influences of linguistic information (arising from multiple levels) on the N400, consistent with contemporary views of the N400 as reflecting activation within a highly distributed, temporally extended, and interactive neural network supporting conceptual processing (Barber & Kutas, 2007; Kutas & Federmeier, 2011). The current results suggest that the incremental influences of both semantic and syntactic context guide semantic processing, as indexed by the N400. Additionally, the findings from the current study indicate that modeling word-level variability in event-related EEG activity can reveal mechanisms by which different sources of information simultaneously contribute to the unfolding neural dynamics of comprehension.

References

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & van der Sluis, S. (2014). A solution to dependency: Using multilevel analysis to accommodate nested data. *Nature Neuroscience*, *17*, 491–496. doi: 10.1038/nn.3648

Ahmadi, M., & Quiroga, R. Q. (2013). Automatic denoising of single-trial evoked potentials. *NeuroImage*, *66*, 672–680. doi: 10.1016/j.neuroimage.2012.10.062

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*, 502–518. doi: 10.1016/j.jml.2006.12.004

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi: 10.1016/j.jml.2007.12.005

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445–459. doi: 10.3758/BF03193014

Barber, H. A., & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, *53*, 98–123. doi: 10.1016/j.brainresrev.2006.07.002

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328. Doi: 10.3389/fpsyg.2013.00328.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. (2015). Parsimonious mixed models. Retrieved June 22, 2015, from arXiv:1506.04967

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, *40*, 373–400. doi: 10.1207/s15327906mbr4003_5

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K. R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage*, *56*, 814–825. doi: 10.1016/j.neuroimage.2010.06.048

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*, 199–231. doi: 10.1214/ss/1009213726

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911. doi: 10.3758/s13428-013-0403-5

Cohen, J., Cohen, P., West, S. G., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Routledge.

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Curran, P. J., Bauer D. J., & Willoughby M. T. (2004). Testing main effects and interactions in latent curve analysis. *Psychological Methods*, *9*, 220–237.

Cumming, G. (2014). The new statistics why and how. *Psychological Science*, *25*, 7–29. doi: 10.1177/0956797613504966

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

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Delorme, A., Miyakoshi, M., Jung, T. P., & Makeig, S. (2014). Grand average ERP-image plotting and statistics: A method for comparing variability in event-related single-trial EEG activities across subjects and conditions. *Journal of Neuroscience Methods*, *250*, 3–6. doi: 10.1016/j.jneumeth.2014.10.003

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

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of Learning and Motivation*, *51*, 1–44. doi: 10.1016/S0079-7421(09)51001-8

Fischer-Baum, S., Dickson, D. S., & Federmeier, K. D. (2014). Frequency and regularity effects in reading are task dependent: Evidence from ERPs. *Language, Cognition and Neuroscience*, *29*, 1342–1355. doi: 10.1080/23273798.2014.927067

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. doi: 10.1016/j.bandl.2014.10.006

Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language*, *74*, 289–300. doi: 10.1006/brln.2000.2313

Gaspar, C. M., Rousselet, G. A., & Pernet, C. R. (2011). Reliability of ERP and single-trial analyses. *NeuroImage*, *58*, 620–629. doi: 10.1016/j.neuroimage.2011.06.052

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. doi: 10.1146/annurev.psych.58.110405.085530

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage*, *17*, 1101–1116. doi: 10.1006/nimg.2002.1268

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford.

Holcomb, P., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, *14*, 938–950. doi: 10.1162/089892902760191153

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. doi: 10.1016/j.jml.2007.11.007

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution*, *5*, 944–946. doi: 10.1111/2041-210X.12225

Kaan, E., & Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences*, *6*, 350–356. doi: 10.1016/S1364-6613(02)01947-2

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

Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *The Quarterly Journal of Experimental Psychology*, *63*, 1838–1857. doi: 10.1080/17470211003602412

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

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1982). The lateral distribution of event-related potentials during sentence processing. *Neuropsychologia*, *20*, 579–590. doi: 10.1016/0028-3932(82)90031-8

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

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified. In *Handbook of Psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press.

Kutas, M., Van Petten, C., & Besson, M, (1988). Event-related potential asymmetries during the reading of sentences, *Electroencephalography and Clinical Neurophysiology*, *69*, 218–233. Doi: 10.1016/0013-4694(88)90131-9

Laszlo, S., & Federmeier, K. D. (2007). Better the DVL you know: Acronyms reveal the contribution of familiarity to single word reading. *Psychological Science*, *18*, 122–126.

Laszlo, S., & Federmeier, K. D. (2008). Minding the PS, queues, and PXQs: Uniformity of semantic processing across multiple stimulus types. *Psychophysiology*, *45*, 458–466. doi: 10.1111/j.1469-8986.2007.00636.x

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

Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*, *48*, 176–186. doi: 10.1111/j.1469-8986.2010.01058.x

Laszlo, S., & Federmeier, K. D. (2014). Never seem to find the time: Evaluating the physiological time course of visual word recognition with regression analysis of single-item event-related potentials. *Language, Cognition and Neuroscience*, *29*, 642–661. doi: 10.1080/01690965.2013.866259

Lazic, S. E. (2010). The problem of pseudoreplication in neuroscientific studies: Is it affecting your analysis? *BMC Neuroscience*, *11*, 1–17. doi: 10.1186/1471-2202-11-5

Lee, C. L., & Federmeier, K. D. (2009). Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. *Journal of Memory and Language*, *61*, 538–555. doi: 10.1016/j.jml.2009.08.003

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*, 707–710.

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.

Locker, L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods*, *39*, 723–730. doi: 10.3758/BF03192962

Luck, S. J. (2014). An introduction to the event-related potential technique. *Cambridge*, MA: MIT Press.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19. doi: 10.1037/1082-989X.7.1.19

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71. doi: 10.1016/0010-0277(80)90015-3

The MathWorks. (2014). *MATLAB, signal processing, and statistics toolboxes*. Natick, MA: Author.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier O., . . . & Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, *5*, 467–479.

Molinaro, N., Conrad, M., Barber, H. A., & Carreiras, M. (2010). On the functional nature of the N400: Contrasting effects related to visual word recognition and contextual semantic integration. *Cognitive Neuroscience*, *1*, 1–7. doi: 10.1080/17588920903373952

Mouraux, A., & Iannetti, G. D. (2008). Across-trial averaging of event-related EEG responses and beyond. *Magnetic Resonance Imaging*, *26*, 1041–1054. doi: 10.1016/j.mri.2008.01.011

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142. doi: 10.1111/j.2041-210x.2012.00261.x

Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G., & Stine-Morrow, E. A. L. (2014). Aging and individual differences in binding during sentence understanding: Evidence from temporary and global syntactic attachment ambiguities. *Cognition*, *130*, 157–173. doi: 10.1016/j.cognition.2013.10.005

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437–448. doi: 10.3102/10769986031004437

Quiroga, R. Q., & Garcia, H. (2003). Single-trial event-related potentials with wavelet denoising. *Clinical Neurophysiology*, *114*, 376–390. doi: 10.1016/S1388-2457(02)00365-6

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457–1506. doi: 10.1080/17470210902816461

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*, 289–310. doi: 10.1214/10-STS330

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *23*, 323–355. doi: 10.2307/1165280

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, UK: Oxford University Press. doi: 10.1093/acprof:oso/9780195152968.001.0001

Smith, N. J., & Kutas, M. (2015a). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*, 157–168. doi: 10.1111/psyp.12317

Smith, N. J., & Kutas, M. (2015b). Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology*, *52*, 169–181. doi: 10.1111/psyp.12320

Snijders, T. A., & Bosker, R. J. (2011). *Multilevel Analysis*. Berlin Heidelberg: Springer. doi: 10.1007/978-3-642-04898-2_387

Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, *52*, 124–139. doi: 10.1111/psyp.12299

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, *8*, 485–531. doi: 10.1080/01690969308407586

Van Petten, C. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology*, *32*, 511–525. doi: 10.1111/j.1469-8986.1995.tb01228.x

Van Petten, C. (2014). Examining the N400 semantic context effect item-by-item: Relationship to corpus-based measures of word co-occurrence. *International Journal of Psychophysiology*, *94*, 407–419. doi: 10.1016/j.ijpsycho.2014.10.012

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, *18*, 380–393. doi: 10.3758/BF03197127

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open-and closed-class words. *Memory & Cognition*, *19*, 95–112. doi: 10.3758/BF03198500

Vergara-Martínez, M., & Swaab, T. Y. (2012). Orthographic neighborhood effects as a function of word frequency: An event-related potential study. *Psychophysiology*, *49*, 1277–1289. doi: 10.1111/j.1469-8986.2012.01410.x

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971–979. doi: 10.3758/PBR.15.5.971

(Received March 6, 2015; Accepted July 26, 2015)

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1: Fitting linear mixed-effects models using the lme4 package in R.

Appendix S2: Best linear unbiased predictors of the random effects.

Appendix S3: Comparative and absolute goodness-of-fit indices.

Appendix S4: Further reading.