Phonological Typicality Influences Sentence Processing in Predictive Contexts: A Reply to Staub et al. (2009)

Thomas A. Farmer

Department of Brain and Cognitive Sciences, University of Rochester

Padraic Monaghan

Department of Psychology, Lancaster University

Jennifer B. Misyak and *Morten H. Christiansen*[*]

Department of Psychology, Cornell University

[*]Please address correspondence to:
Morten H. Christiansen
Department of Psychology
228 Uris Hall
Cornell University
Ithaca, NY 14853
Phone: 607-255-3834
e-mail: christiansen@cornell.edu

Abstract

In two separate self-paced reading experiments, Farmer, Christiansen and Monaghan (2006) found that the degree to which a word's phonology is typical of other words in its lexical category influences on-line processing of nouns and verbs in predictive contexts. Staub, Grant, Clifton and Rayner (2009) failed to find an effect of phonological typicality when they combined stimuli from the separate experiments into a single experiment. We replicated Staub et al.'s experiment and found that the combination of stimulus sets affects the predictiveness of the syntactic context; this reduces the phonological typicality effect as the experiment proceeds, though the phonological typicality effect was still evident early in the experiment. Although an ambiguous context may diminish sensitivity to the probabilistic relationship between the sound of a word and its lexical category, phonological typicality does influence on-line sentence processing during normal reading when the syntactic context is predictive of the lexical category of upcoming words.

Phonological Typicality Influences Sentence Processing in Predictive Contexts: A Reply to Staub et al. (2009)

Language comprehension is a complex task that involves constructing an incremental interpretation of a rapid sequence of incoming words before they fade from immediate memory, and yet the task is typically carried out efficiently and with little conscious effort. In order to achieve this level of speed and efficiency, the adult comprehension system exploits multiple sources of information that might facilitate the task. Many factors, including referential context (e.g., Altmann, Garnham, & Dennis, 1992; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002), lexically-based verb biases (e.g., Trueswell, Tanenhaus, & Kello, 1993), plausibility (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997) and prosody (e.g., Snedeker & Yuan, 2008) appear to constrain how an incoming string of words is processed (see Altmann, 1998; Elman, Hare, & McRae, 2004, for reviews). Such informative cues are not only used to resolve previously encountered ambiguous input, but also to generate syntactic expectations for what may come next. Indeed, a growing number of studies suggest that prediction-based processing is a necessary component of efficient and effortless interpretation of language as it unfolds in time (e.g., Altmann, 1998; Rayner, Ashby, Pollatsek, & Reichle, 2004; Staub & Clifton, 2006; see Hagoort, 2009; Pickering & Garrod, 2007, for reviews).

Convergent results have been found in event-related potential (ERP) experiments (see Federmeier, 2007, for a review), showing that highly specific expectations are generated for both lexical-category and phonological properties of upcoming words given a predictive context. Thus, during on-line sentence processing, context-based expectations are rapidly generated for (a) the grammatical gender of upcoming words, such as specific gender markings of nouns following a gender-marked adjective in spoken Dutch (Van Berkum, Brown, Zwitserlood,

Kooijman, & Hagoort, 2005) or a gender-marked adjective in written Spanish (Wicha, Moreno, & Kutas, 2004), (b) the lexical category of the next word (e.g., a noun following a determiner, Hinojosa, Moreno, Casado, Muñoz, & Pozo, 2005), and (c) the onset phoneme of the next word (e.g., words starting with a consonant after '*a*' or a vowel after '*an*' in English, DeLong, Dubach, & Kutas, 2005).

Building on this work, Farmer, Christiansen, and Monaghan (2006) investigated whether phonological typicality—the degree to which the sound properties of an individual word are typical of other words in its lexical category—influences on-line language processing in predictive contexts, testing a hypothesis originally put forward by Kelly (1992) and supported by recent work on language acquisition (e.g., Cassidy & Kelly, 2001; Fitneva, Christiansen, & Monaghan, 2009; Monaghan, Christiansen, & Chater, 2007). Farmer et al. presented results from a corpus analysis, showing that nouns tend to sound like other nouns and verbs like other verbs; that is, nouns and verbs form separate coherent, yet partially overlapping, clusters in phonological space. Thus, some words are more typical in their phonology of their respective lexical class than others. Farmer et al. referred to words that are typical, in terms of their phonology, of the class of nouns as 'noun-like,' and words more phonologically typical of verbs as 'verb-like'. They then reported four experiments demonstrating the impact of such phonological typicality on the processing of nouns and verbs. Using a self-paced reading methodology, two of the experiments focused on the processing of unambiguous sentences and elicited significant effects of phonological typicality. One experiment involved sentence frames designed to strongly predict that a noun will come next, whereas the frames in the other experiment were created to generate strong expectations for a verb. When the preceding context generated a strong expectation for an upcoming noun, noun-like nouns were read faster than

verb-like nouns, and when the context was highly predictive of a verb, verb-like verbs were read faster than noun-like verbs.

Tanenhaus and Hare (2007) noted that studies of eye-movement patterns during reading have found that initial fixation durations on words are relatively uninfluenced by various types of higher-level linguistic information (e.g., plausibility, referential context, and so forth) that typically exert an influence on later processing. They argued that during reading, it is possible that predictions about upcoming word forms are being generated, and that various cues to word form, such as phonological typicality, may be the types of factors that would influence indices of early processing such as the duration of initial fixations. This hypothesis was confirmed by Dikker, Rabagliati, Farmer, and Pylkkanen (2010). Using magnetoencephalography (MEG), Dikker et al. demonstrated that the visual M100 response, a component in visual cortex that arises approximately 100-130 milliseconds (ms) after stimulus onset in response to sensory-based violations of expectations while reading (Dikker, Rabagliati, & Pylkkanen, 2009), is sensitive to phonological typicality. They found that an effect of expectedness of a noun (should a noun be next or not) was modulated by the phonological typicality of the incoming noun. In a condition where all nouns had phonological properties highly typical of nouns, the effect of expectedness was larger than in a condition where all of the nouns were neutral in terms of their phonology. That is, the magnitude of the M100 was significantly larger when a noun was not expected but nonetheless occurred and was highly typical of other nouns in terms of its word form, compared to when a noun was expected. When the nouns were not typical or atypical of other nouns (neutral), there was no difference in M100 magnitude in the expected versus the unexpected condition. This effect appears to be generated in the visual cortex while reading, and is in-line with the Tanenhaus and Hare proposal (also advanced in Dikker et al, in 2010) that

while reading, word-form predictions of upcoming material are being generated and available to the visual cortex. Nonetheless, it accentuates the role that word-form predictions play during language processing, along with the importance of a highly constraining (or predictive) preceding sentential context for producing an effect of phonological typicality.

Recently, Staub, Grant, Clifton and Rayner (2009) failed to find effects of phonological typicality in experiments examining eye-tracking and self-paced reading times when they combined the unambiguous noun and verb materials from Farmer et al.'s (2006) two separate experiments. Staub et al. interpreted their null results as indicating that phonological typicality may not influence normal reading. In the study that follows, we demonstrate that the replication failure may be due to an unforeseen consequence of Staub et al.'s interleaved design, and that when this design characteristic is accounted for, the effect of phonological typicality re-emerges.

Consider the following examples of the experimental sentences from Farmer et al. (2006), and used in Staub et al. (2009):

(1a) The curious young boy saved the *marble* that he … (Noun-like Noun)

(1b) The curious young boy saved the *insect* that he … (Verb-like Noun)

(2a) The very old man attempted to *assist* his elderly wife ... (Verb-like Verb)

(2b) The very old man attempted to *vary* his daily routine ... (Noun-like Verb)

As illustrated in (3), there is little difference in sentence structure between the noun (1) and verb (2) items up until the word following the main verb of each sentence frame:

(3) NP V the/to <*critical N/V*>...

The main verbs were strongly biased to generate expectations for a NP for the noun items, and for an infinitival complement for the verb items (see Farmer et al., for information about these biases). The critical nouns and verbs may be predicted by the immediately preceding function

word, ‘the’ or ‘to’. However, up to that point, there is a complete overlap of syntactic material for both noun and verb items: both begin with a NP followed by a V. We therefore contend that predictive context is likely to accumulate throughout the overlapping sentence frame, and is not dependent only on the function word preceding the critical noun or verb. When these stimuli are intermixed, the extent of this overlap is likely to reduce the distinctiveness between critical-noun and critical-verb sentence stimuli. At the beginning of the experiment, this information may assist in biasing the participant toward a particular reading, but with repeated instances of this structure, the participant may learn that an initial NP followed by a V does not provide a reliable indication of upcoming syntactic structure, therefore reducing the biasing context for the critical-nouns and critical-verbs as the experiment proceeds. Accordingly, at the onset of the experiment, the participant may be using the entire sentence frame to predict the category of the target word. By the end of the experiment, the participant has learned to disregard most of the frame as predictive of category.

Stated alternatively, the word order common to the beginnings of the experimental items may be acting as another cue to structure. Early in the study, the verb bias acts alone as a strong cue to whether a noun or verb is likely to occur next. However, as subjects progress through the study, they are likely to pick up on the commonality of the sentence-initial structure, and the fact that the structure can be continued with a noun or verb. Given the large amount of literature on the ease with which children and adults can map regularities that are often subtle in nature during artificial language learning tasks (e.g., Perruchet & Pacton, 2006; Pothos, 2007), it is likely that subjects implicitly learn to recognize the structure shared between the N and V items in the interleaved design, and that when such a word order is used, the main verb can be followed by either an N- or V-structure. The net effect is that once subjects learn that the structure of the

preamble is common to a set of items in which a main verb can be followed by either a N or V content word, the strong effect of the verb bias for forcing an expectation for a N- or V-structure becomes a less reliable cue over the course of the experiment. This reduction in predictiveness of the grammatical category of the word, then, is a consequence of the experimental manipulation. Contextual predictiveness, which is a property of natural language (see, e.g., Federmeier, 2007; Pickering & Garrod, 2007, for reviews), may therefore be weakened in the Staub et al. (2009) study.

The hypothesized decrease of the main verb biases in the noun and verb items over the course of the experiment amounts to a learning effect. The effects of such learning during traditional sentence processing experiments are not currently well understood (but see Fine, Qian, Jaeger, & Jacobs, 2010). Although traditional statistical analyses such as regression or Analysis of Covariance (ANCOVA) could feasibly be used to investigate how the influence of an independent variable may change with repeated exposure to the critical regions of sentences containing manipulations of that variable, they have rarely been employed with such a goal in mind. As Baayen, Davidson, and Bates (2008) have noted, however, the linear fixed effects modeling approach utilized by Staub et al. (2009) is particularly well-suited to illuminate the manner in which particular effects may change across the course of an experiment. Here, we exploit this advantage in order to demonstrate that subject responses to the experimental items did indeed change during the experiment.

In the study presented next, we followed Staub et al. (2009) in combining the original noun and verb items from Farmer et al.'s (2006) two separate experiments within a single self-paced reading experiment. If combining items that produce a strong expectation for a noun with the items that produce a strong expectation for a verb reduces the context-driven prediction for

target words of either lexical category as the experiment progresses, we should make two observations:

1) When conducting the same linear mixed-effects analysis that Staub et al. report in their Experiment 2 (on self-paced reading), we should replicate their lack of a significant interaction between Part of Speech (PoS) and Phonological Classification (PC; whether the target word is Noun-like or Verb-like).

2) When adding Presentation Order to the model as a fixed effect, allowing it to interact with PoS and PC, we should observe a PoS x PC x Order interaction. The phonological typicality effect—noun-like nouns being read faster than verb-like nouns in the noun context, and verb-like verbs being read faster in the verb context—should be present for the items that subjects encountered early in the experiment, when the biases exerted by the initial sentential context remain strong due to the fact that subjects have not had the opportunity to learn about the regularities associated with the experimental items. Later in the experiment, when expectations for either a noun or a verb have been attenuated, the typicality effect should weaken.

Method

Participants

Forty undergraduate native English speakers from Cornell University (M=19.54 years, SD=1.10) participated for extra credit in a psychology course.

Materials

For both the noun and verb items, two sentence versions were constructed from each sentence frame. One version included a noun phrase with a noun-like noun (*marble,* 1a), and the

other version contained a verb-like noun (*insect,* 1b). For the verb items, one version of each sentence frame contained an infinitival complement containing a verb-like verb (*assist*, 2a), and the other version contained a noun-like verb (*vary*, 2b). For both the noun and verb items, there was no significant difference in CELEX- and HAL-based lexical frequency, orthographic length, number of phonemes, number of phonological neighbors (also from CELEX), or plausibility (obtained from plausibility norming studies on separate groups of subjects—originally reported in Farmer et al., 2006, pp. 12207-12208) between the phonologically typical versus atypical items. The 20 experimental items (10 noun and 10 verb items) were combined and then counterbalanced across two different presentation lists in such a way that each list contained five noun-like noun sentences, five verb-like noun sentences, five verb-like verb sentences, and five noun-like verb sentences, but only one version of each of the 20 frames. Each list also contained 30 unrelated filler items and eight practice items. A majority of the filler sentences contained reduced or unreduced relative clauses, and the others were simple unambiguous sentences containing no relevant psycholinguistic manipulations.

Procedure

Subjects were randomly assigned to one of the two presentation lists. The order in which all items contained in each presentation list, either filler or experimental, were presented was randomized separately for each subject. All sentences were presented in a non-cumulative, word-by-word moving window format using PsyScope version 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993). After a brief tutorial, subjects were instructed to press the 'GO' key to begin the task. For all sentences, the entire test item appeared left-justified at the vertical center of the screen in such a way that dashes preserved the spatial layout of the sentence, but masked the actual characters of each word. As the subjects pressed the 'GO' key, the word that was just read

disappeared and the next one appeared. RTs (ms) were recorded for each word. After each sentence had been read, subjects responded to a Yes/No comprehension question, and upon another key press, the next item appeared.

Results and Discussion

One participant reported the presence of an auditory processing deficit and was excluded from all subsequent analyses. Overall accuracy on the comprehension questions relating to the 20 experimental sentences was close to ceiling (M=19.44 correct, SD=1.14), and no significant main effect of PC or PoS, or interaction, was observed on accuracy rates, all F's < 1.3. In keeping with the original Farmer et al. experiments, the focus of our analyses was on the critical word that contained the experimental manipulation of phonological typicality. All RTs over 2000 ms were excluded from the subsequent analyses, resulting in the omission of five trials (less than 1% of the data).

The mean RTs on the critical word for each condition are presented in Figure 1. The means for the typical words are slightly lower than the means for the atypical words in both the noun and verb conditions. As in Staub et al. (2009), RTs on the critical word were analyzed in a linear mixed-effects model using the lme4 package in R[1] (R Development Core Team, 2007), and the analyses will be presented twice, first without the inclusion of presentation Order, as in Staub et al.'s analysis, and second with Order as an additional fixed factor. Order was coded by labeling the experimental items that subjects saw with a number between 1 and 20, reflecting the order in which each experimental item was viewed by each subject. In the first analysis (not considering potential effects of order), RTs were the dependent measure, subjects and items were

[1] We are grateful to Adrian Staub and Margaret Grant for making the R syntax for their statistical analyses available to us.

entered as crossed random factors, and the fixed factors were PoS, PC, the PoS x PC interaction, length, and HAL-based log frequency. All parameter estimates, as well as *p*-values (estimated by Markov Chain Monte Carlo sampling, Baayen, 2008) associated with the t-tests for each effect, are listed in Table 1. As is evident in Table 1, the results were similar to those of Staub et al. in that there was no significant effect of PoS or PC, no significant interaction between PoS and PC, and no significant effect of frequency. Unlike Staub et al., however, there was a significant effect of length in the present dataset, with longer words being read more slowly.

In order to assess the hypothesis that the effect of phonological typicality would diminish as the experiment progressed, we conducted the same analysis detailed above, except that presentation Order was entered as a fixed effect, interacting with PoS and PC. Table 2 displays the parameter estimates and p-values associated with each term in the model. The effect of Order, by itself, was not significant, and did not interact with PoS. The three-way interaction between Order, PoS, and PC, $p = .046$, indicated that the interaction between PoS and PC was dependent upon Order.

To illustrate the influence of presentation Order on the phonological typicality effect, bins of items were generated based on whether the items of each PoS condition appeared early or late in the experiment for each participant. More specifically, one bin contained the first five noun items encountered by each participant, and another contained the last five noun items. Bins were also created for the first and last five verb items. Note that this was not the same as analyzing the first and last ten sentences in the experiment, as order was randomized for each subject. Additionally, to measure the extent to which the syntactic expectancies for a NP or infinitival complement faded as the experiment progressed—thus diminishing the typicality effect—we also generated bins for the first and last three noun and verb items. Then, within both

the early and late bins for each PoS, the magnitude of the typicality effect was graphically assessed.

Figure 2 shows the predicted effect of Order for the verb items. For both the first/last-five and the first/last-three verb items, verb-like verbs were read more quickly than noun-like verbs at the beginning of the experiment, but in the latter portion of the experiment, the effect of PC disappeared. As illustrated in Figure 3, there is a similar pattern for the noun items. The typicality effect existed, in the predicted direction, for the early items. Interestingly, however, the typicality effect was larger for the first three items compared to the first five. The pattern of effects differ somewhat for the final noun and verb items, suggesting that predictiveness of prior context may affect noun and verb phonological typicality in slightly different ways. In this case, context-driven expectancies appear to influence nouns more than verbs, perhaps because phonological typicality may be a stronger factor for verbs than for nouns. In corpus-based research, for example, Christiansen and Monaghan (2006) found that phonological information provides a better cue to verbs whereas distributional information is more likely to affect the learning and processing of nouns. Similarly, Fitneva et al. (2009) elicited stronger phonological typicality effects for verb-like than for noun-like non-words.

Based on the pattern of mean RTs depicted in Figures 2 and 3, it may be objected that the significant three-way interaction could be explained by a reversal of the condition means (atypical words being read more quickly than typical words) at the end of the study, as opposed to the effect existing in the predicted direction at the beginning of the study. Follow-up tests do not, however, offer support for this suggestion. *T*-tests on the RTs between the typical and atypical conditions for each lexical category were not significant for either the early- or late-occurring item bins (although when examining the 3-item bins, a one-tailed *t*-test on the

difference between the noun-like and verb-like nouns in the first-three-item bins was near-significant in the predicted direction, $p = .07$). Additionally, the two-way PoS x PC interaction was not significant in the early or late bins across items from each lexical category. Investigation of the mean differences, however, revealed that across each lexical category, the mean difference between typical and atypical conditions was larger, and in the predicted direction, at the early-item bins than it was in the late-item bins. Indeed, for the verb items, there was basically no reverse effect at all in the final-item bins. This supports the notion that the three-way interaction is driven by the phonological typicality effect existing in the predicted direction at the beginning of the experiment rather than the more slight effect in the opposite direction in the late-item bins.

These analyses thus provide an explanation for Staub et al.'s (2009) failure to replicate the results from two of the original experiments reported in Farmer et al. (2006) in terms of learning effects that weaken sentential context. However, there are additional contributory factors to the weakened effect of phonological typicality in Staub et al.'s study. In their first experiment, they included filler items that were "designed to determine subjects' interpretation of ambiguous or semantically odd sentences" (p. 808). As one example, some filler sentences included words that were semantically incongruent with their corresponding sentence contexts, such as "The man used the phone to call the old frame together." Although it is unclear what effect the presence of "ambiguous or semantically odd sentences" can have on the processing of well-formed sentences within a single experiment, previous research has demonstrated, for example, that the ratio of grammatical to ungrammatical filler items can influence the degree to which effects are elicited by linguistic manipulations (e.g., Hahne & Friederici, 1999). Thus, this deviation from the original experimental design may also have had repercussions for the types of effects originally reported by Farmer et al.

Additionally, in each of their experiments, Staub et al. (2009) created new sentence frames so that subjects would be exposed to both the typical and atypical words from each of the original items. Instead of having two versions of one sentence frame (one containing a typical and other containing an atypical word), Staub et al.'s subjects saw either the typical word in its original frame, and the corresponding atypical words in a newly created frame, or vice versa. Although they argue that this modification increased the power of the study (thus making it easier to observe an effect should one be present), it turns out, upon examining Staub et al.'s newly created frames, that these are, in some cases, semantically minimally different to the original frames. The beginning of the newly created frame for 1b, for example, is "The retired man attempted …" instead of the original "The very old man attempted…" The fact that for each of our original typical-atypical item-pairings that were counterbalanced across two presentation lists, a subject saw both words appearing in a highly similar semantic and syntactic context raises the possibility that responses to the second-occurring word in the item-pair are influenced by the presence of a word of the opposite phonological typicality valence appearing before it.

General Discussion

In our replication of Staub et al.'s (2009) study, we found that phonological typicality is influenced by learning effects deriving from changes in syntactic expectancies as a consequence of the experimental context. Although Staub et al. report a failure to replicate an interaction between Part of Speech and Phonological Classification, it must be noted that no such interaction was reported in Farmer et al. (2006). The original unambiguous noun and unambiguous verb experiments were conducted separately in order that phonological typicality effects could be observed in contexts where the sentence frame was predictive of a particular grammatical

category at the point of interest in the sentence. Based on their data, Staub et al. prematurely claimed that the phonological typicality effects reported in Experiments 2 and 3 of Farmer et al. were likely the result of a Type I error. Instead, the data presented here suggest that Staub et al.'s null results may be traced to their altering of the original Farmer et al. design by interleaving syntactic frames that generate a strong expectation for a noun with those that are highly predictive of verbs. Using their interleaved design, we found that without accounting for Order, there was no significant interaction between PoS and PC. However, including the three-way interaction between PoS, PC, and Order, it becomes apparent that Order influenced the nature of the interaction between PoS and PC. The effects of presentation Order observed here provide support for our hypothesis that the overlap in syntactic context preceding the critical words is contributing to the reduction of the strength of the expectation for either a noun or a verb over time, with a negative impact on the phonological typicality effect. As predicted by this hypothesis, we found that the typicality effect for each grammatical category decreased as the experiment progressed. For both the noun and verb items, the phonological typicality effect was observed for the items presented early, where main verb biases from natural language situations for either a NP or VP would be strongest, and was attenuated across the course of the experiment.

The interpretation of the data from the interleaved design offered here may seem, at face value, problematic when considered in conjunction with the results of Dikker et al. (2010). In their experiment, item-types were intermixed, and still, an effect of phonological typicality was observed. It is important to note, however, that in the experiment detailed there, only responses to nouns were studied, and the linguistic manipulation differed substantially from the one reported here. Nouns that varied in their degree of nouniness (either very noun-like, or neutral, in

terms of their phonological typicality scores) were shown to subjects in both sentence-initial predictive (*The tasteless **soda***) or non-predictive (*The tastelessly **soda***) contexts. Unlike the study presented here, subjects saw the target words multiple times in an equal number of predictive and non-predictive contexts. Because the manipulation always occurred at the beginning of a sentence, directly after the determiner "The," no pre-critical-region syntactic cues existed to facilitate a prediction about word-category information.

Staub et al. (2009) suggested that should intermixing the noun and verb items cause the elimination of the phonological typicality effect, then the effect would "reflect task-dependent strategic factors as opposed to the processes involved in normal word recognition" (p. 813). In contrast, the fact that the phonological typicality effect is observed *early* in the experiment indicates that phonological typicality exerts its effect *before* any potential strategic effects would be likely to occur. Participants have expectations of contexts derived from experience with natural language, that we probed in our norming studies in Farmer et al. (2006). However, during the course of the interleaved experiment the contextual expectations from natural language appear to be weakened, and consequently the effects of potential cues to the lexical category of the upcoming word are less likely to be observed. This hypothesis about effects of weakened context in the interleaved experimental design has as a corollary that there should be no effect of order in the original blocked design studies of Farmer et al., as the predictive context of natural language is maintained throughout the blocked design experiments[2]. In linear mixed effects analyses of the noun and verb blocked studies (Experiments 2 and 3 of Farmer et al., 2006), Order did not interact significantly with PC, $p = .884$ and $p = .191$, respectively). More generally, the effect of the experimental context on sentence processing, as revealed by the effect

[2] We thank an anonymous reviewer for suggesting these analyses.

of Order using the linear mixed effects analysis, opens up intriguing possibilities for exploring effects of natural language context early in an experiment, as well as learning effects within a study as the experiment proceeds.

The effect of learning during an experiment is something that sentence processing researchers know little about. In a traditional sentence processing experiment, multiple versions of a single sentence frame are created, each containing some different level of a linguistic variable of interest. The different versions of each item are then carefully counterbalanced across a series of presentation lists so that subjects see only one version of each item. To help ensure that participants do not catch on to the manipulation of interest, a series of "filler" items are intermixed with the experimental items in each presentation list. The problem is, though, that even if filler items help to prevent subjects from noticing the actual experimental manipulation, it is still the case that within one presentation list, there exists a subset of items to which subjects are exposed that tend to have a large amount of structural (and often times semantic) overlap among them (as with our items in the interleaved design, the structure and focus of the sentence up until the point where the manipulation occurs, are highly overlapping). In certain cases, the semantic and structural overlap among a subset of items may exert an influence on patterns of processing that have unintended consequences for the interpretation of the behavior elicited by the linguistic stimuli.

Consistent with Tanenhaus and Hare's (2007) view and from the data contained in Dikker et al. (2010), phonological typicality is likely to be one of many word-form cues that are exploited during the early part of language processing in order to facilitate the interpretation of the incoming signal. When words are presented in isolation, an effect of phonological typicality has been observed across different psycholinguistic tasks. For example, in a word learning study,

children were guided by phonological typicality when asked to match noun-like and verb-like nonwords to pictures of actions and objects (Fitneva et al., 2009). In addition, Monaghan, Christiansen, Farmer, and Fitneva (in press) found that although phonological typicality effects may be small, they are nonetheless robustly observed for naming and lexical decision response times for nouns and verbs across a variety of different operationalizations of phonological typicality. When nouns and verbs were read in sentential contexts strongly predictive of their respective lexical category, Farmer et al. (2006) also obtained significant effects of phonological typicality. However, when the surrounding syntactic context is not as reliable, other word form cues that are probabilistically related to lexical category may usurp the usefulness of phonological typicality for processing. As we have shown in the three-way analysis with Order, such effects are subtle, complex, and highly interactive. Thus, we do not see the results reported here as an end-point, but rather as a launching pad for further experimental investigations into the relationship between phonological typicality, syntactic context, and other variables known to influence normal reading, especially during the earlier moments of real-time language processing.

References

Altmann, G. T. M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Science, 2*, 145-152.

Altmann, G. T., Garnham, A., & Dennis, Y. (1992). Avoiding the garden-path: Eye movements in context. *Journal of Memory and Language, 31*, 685-712.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed effects modeling with crossed random factors for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin and Review, 8*, 519-523.

Christiansen, M. H. & Monaghan, P. (2006). Discovering verbs through multiple-cue integration. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets words: How children learn verbs* (pp. 88-107). New York: Oxford University Press.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1992). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257-271.

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.

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkanen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science, 21*, 629-634.

Dikker, S., Rabagliati, H., & Pylkkanen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition, 110*, 293-321.

Elman, J. L., Hare, M., & McRae, K. (2004). Cues, constraints, and competition in sentence processing. In M. Tomasello & D. I. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates* (pp. 111-138). Mahwah, NJ: Erlbaum.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*, 12203-12208.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*, 491-505.

Fine, A. B., Qian, T., Jaeger, T. F., & Jacobs, R. (2010). Syntactic adaptation in language comprehension. In *Proceedings of ACL Workshop on Cognitive Modeling and Computationa Linguistics*. Uppsala, Sweden.

Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language, 36*, 967-997.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37,* 58-93.

Hagoort, P. (2009). Reflections on the neurobiology of syntax. In D. Bickerton & E. Szathmáry (Eds.), *Biological foundations and origin of syntax. Strüngmann Forum Reports, Vol. 3* (pp. 279-296). Cambridge, MA: MIT Press.

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

Hinojosa, J. A., Moreno, E. M., Casado, P., Muñoz, F., & Pozo, M. A. (2005). Syntactic expectancy: An event-related potentials study. *Neuroscience Letters, 378*, 34-39.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99*, 349-364.

Monaghan, P., Christiansen, M. H., Farmer, T. A., & Fitneva, S. A. (in press). Measures of phonological typicality: Robust coherence and psychological validity. *The Mental Lexicon*.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*, 259-305.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233-238.

Pickering. M. J. & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105-110.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133*, 227-244.

R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 720-732.

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 58,* 574-608.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447-481.

Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: Evidence from Either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 425-436.

Staub, A., Grant, M., Clifton, C., Jr., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 806-814.

Tanenhaus, M. K., & Hare, M. (2007). Phonological typicality and sentence processing. *Trends in Cognitive Science, 11*, 93-95.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 528-553.

Van Berkum, J. J., 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*, 443-467.

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.

Acknowledgments

This work was supported by a Dolores Zohrab Liebmann fellowship awarded to the first author. We would like to thank Mateo Obregon at the University of Edinburgh and Marc Brysbaert at the Universiteit Gent for assistance with the analyses presented here. Thanks are also due to Alex Fine at the University of Rochester for helpful discussions about learning effects in sentence processing experiments and Suzanne Dikker for her insights into the effects reported here.

Table 1. Parameter estimates (and 95% Confidence Intervals) for the mixed-effects model on critical-word RTs without including the effect of presentation Order.

	Estimate	95% CI of Estimate	*p*-value
Intercept	329.86	(189.76, 471.12)	.0001
Part of Speech (PoS)	-15.15	(-65.90, 32.56)	.537
Phonological Classification (PC)	16.68	(-21.50, 56.63)	.392
PoS x PC	-32.43	(-87.83, 22.52)	.246
Length	20.37	(5.19, 35.88)	.010
Log Frequency	-4.59	(-18.63, 7.23)	.464

Table 2. Parameter estimates (and 95% Confidence Intervals) for the mixed-effects model on critical-word RTs, including presentation Order as a fixed effect.

	Estimate	95% CI of Estimate	*p*-value
Intercept	364.03	(17.68, 516.69)	.0001
Part of Speech (PoS)	31.93	(-58.25, 114.14)	.466
Phonological Classification (PC)	94.81	(13.47, 178.91)	.024
PoS x PC	-130.61	(249.36, -22.10)	.028
Length	20.49	(5.79, 36.33)	.009
Log Frequency	-6.63	(-19.53, 5.83)	.284
Order	-1.46	(-6.29, 3.44)	.557
PoS x Order	-4.62	(-11.23, 2.14)	.174
PC x Order	-7.78	(-14.60, -.82)	.028
PoS x PC x Order	9.62	(.32, 19.27)	.046

Figure Captions

Figure 1. Mean RTs on the critical word for each condition of the PoS x PC interaction. Error bars represent standard error of the mean.

Figure 2. Mean RTs across the first-and-last five (left) and three (right) verb items. Error bars represent standard error of the mean.

Figure 3. Mean RTs across the first-and-last five (left) and three (right) noun items. Error bars represent standard error of the mean.

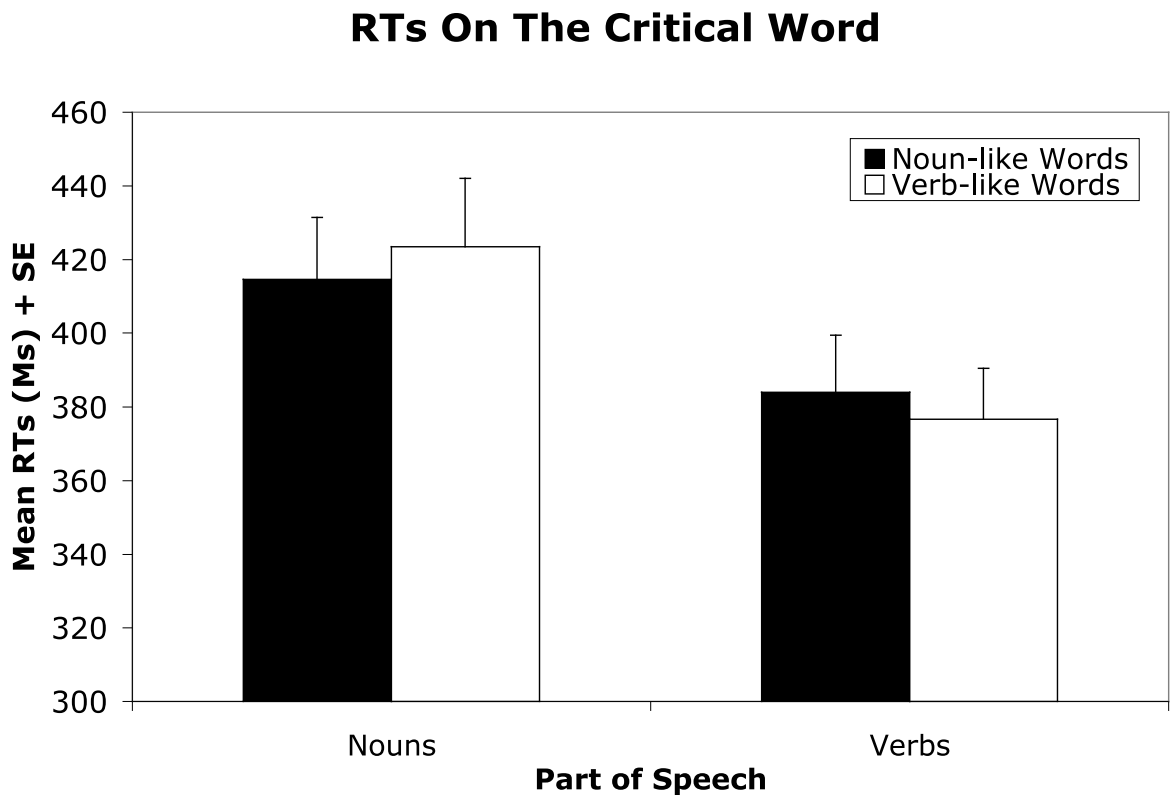
RTs On The Critical Word
460
440
420
400
380
360
340
320
300
Mean RTs (Ms) + SE
Noun-like Words
Verb-like Words
Nouns
Verbs
Part of Speech

Figure 1.

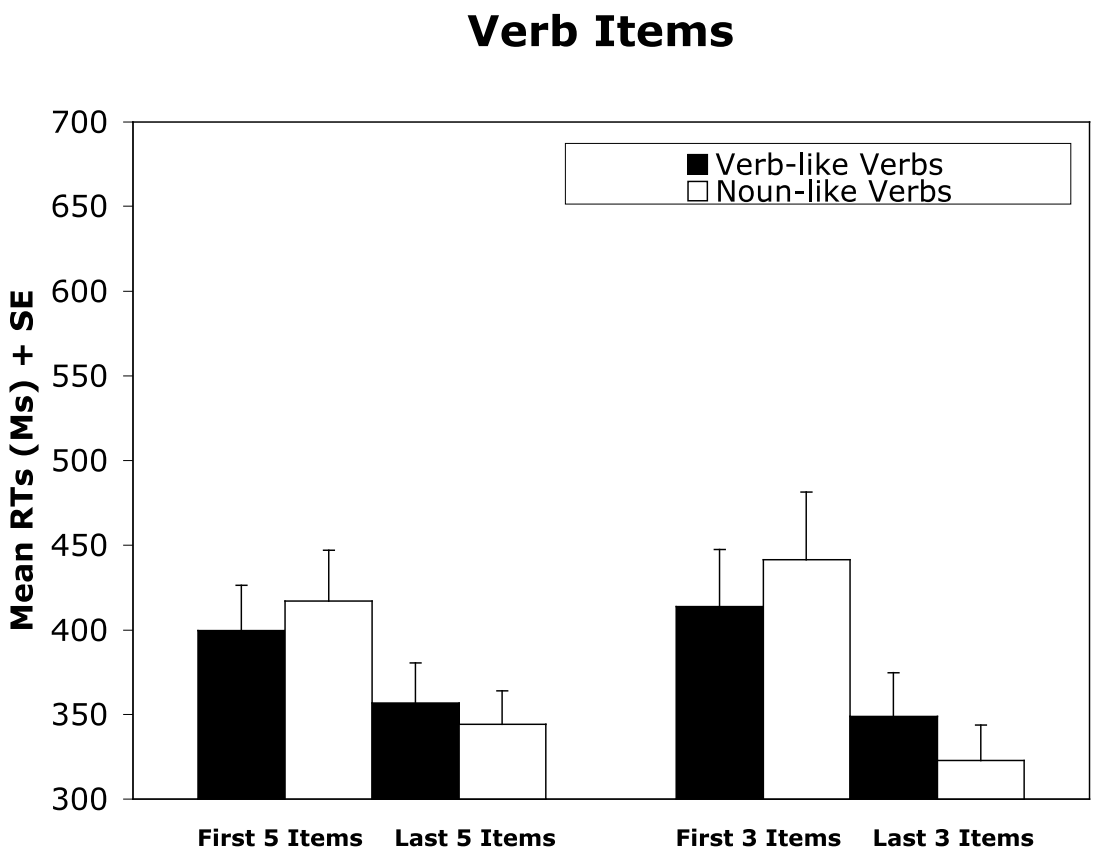
Verb Items
Verb-like Verbs
Noun-like Verbs
Mean RTs (Ms) + SE
700
650
600
550
500
450
400
350
300
First 5 Items
Last 5 Items
First 3 Items
Last 3 Items

Figure 2.

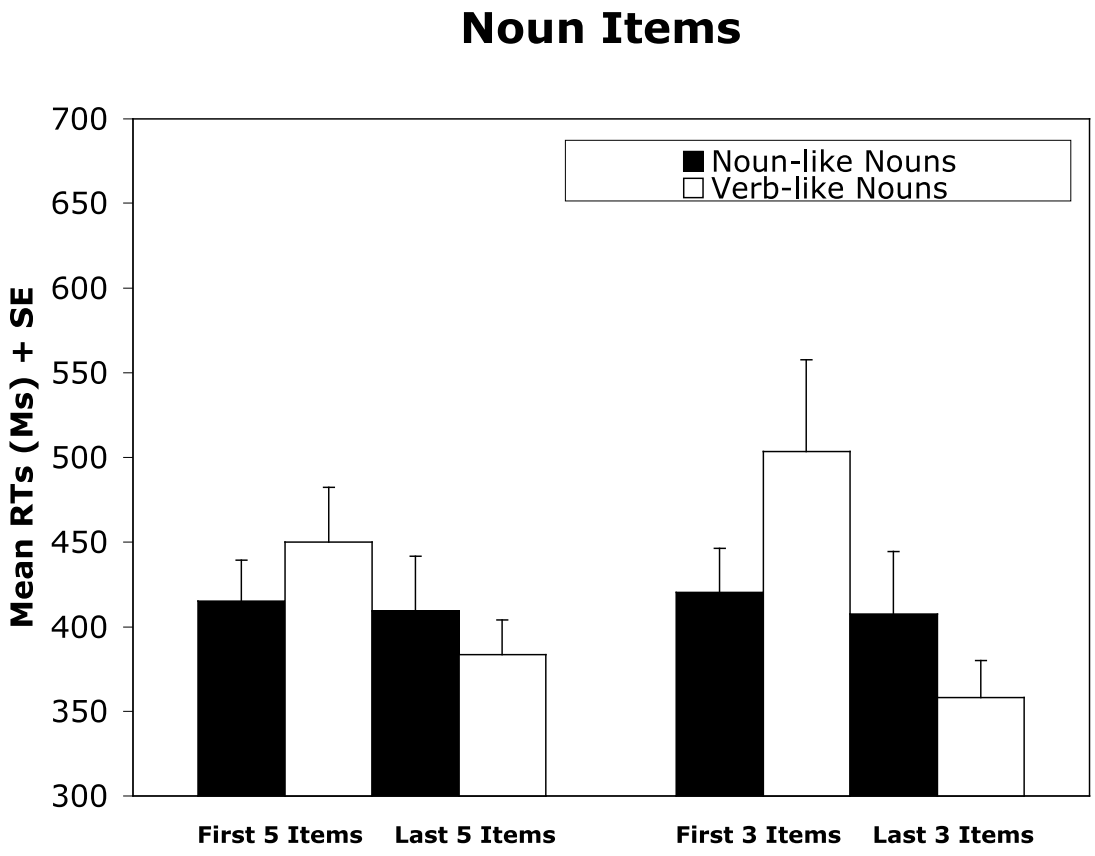
Noun Items
Noun-like Nouns
Verb-like Nouns
Mean RTs (Ms) + SE
700
650
600
550
500
450
400
350
300
First 5 Items
Last 5 Items
First 3 Items
Last 3 Items

Figure 3.