HHS Public Access
Author manuscript
*J Mem Lang*. Author manuscript; available in PMC 2016 February 16.

Published in final edited form as:
*J Mem Lang*. 2015 ; 79-80: 18–29. doi:10.1016/j.jml.2014.11.001.

# Identifying the role of phonology in sentence-level reading

**Dave Kush**, **Clinton L. Johns**, and **Julie A. Van Dyke**
Haskins Laboratories

## Abstract

Phonological properties of the words in a sentence have been shown to affect processing fluency and comprehension. However, the exact role of phonology in sentence comprehension remains unclear. If constituents are stored in working memory during routine processing and accessed through their phonological code, phonological information may exert a pervasive influence on post-lexical comprehension processes such as retrieval for thematic integration. On the other hand, if access to constituents in memory during parsing is guided primarily by syntactic and semantic information, the parser should be isolated from phonologically based effects. In two self-paced reading experiments, we tested whether phonological overlap between distractors and a retrieval target caused retrieval interference during thematic integration. We found that phonological overlap creates difficulty during the initial encoding of the filler, but there was no evidence that phonological overlap caused later interference when the filler was retrieved for thematic integration. Despite effects at encoding, phonological interference did not have a detrimental effect on comprehension. These results suggest that phonological information is not used as a retrieval cue during routine dependency construction in incremental sentence processing. We conclude by considering the potential importance of phonology in parsing under conditions of extraordinary syntactic and/or semantic interference.

### Keywords

sentence processing; sentence comprehension; retrieval interference; phonological interference; working memory

---

Previous research revealing detrimental effects of phonological overlap among sentence constituents suggests that phonology may play a role in sentence comprehension. Specifically, participants' overall reading speed, and/or comprehension accuracy, can be impaired if a sentence contains phonologically similar constituents relative to sentences with identical syntactic structures whose constituents do not have phonological overlap (Acheson & MacDonald, 2011; Baddeley, Eldridge, & Lewis, 1981; Baddeley & Hitch, 1974; Haber & Haber, 1982; Keller, Carpenter, & Just, 2003; Kennison 2004; Kennison, Sieck, & Briesch, 2003; McCutchen, Bell, France, & Perfetti, 1991; McCutchen, Dibble, & Blount, 1994; McCutchen & Perfetti, 1982; Robinson & Katayama, 1997; Zhang & Perfetti, 1993). However, in spite of this empirical record, the mechanism by which phonological overlap causes processing difficulty is uncertain. Furthermore, little is known about the time-course

Correspondence should be addressed to: Dave Kush, Haskins Laboratories, 300 George St.New Haven, CT 06511, kush@haskins.yale.edu.

of these effects during incremental sentence processing because most previous studies employed methods that do not provide fine-grained temporal information. To our knowledge, there is only a single study (Acheson & MacDonald, 2011) that provides evidence that bears on the question of time course. In this study, processing difficulty arose *immediately after* encountering the first phonologically overlapping constituent. That is, despite having identical syntactic structures, reading times were slower for the three words that followed *banker* in phonologically overlapping sentences like (1a), as compared to those like (1b). Comprehension accuracy was also lower in overlapping conditions.

(1a) The baker that the banker sought bought the house.

(1b) The runner that the banker feared bought the house.

Acheson and MacDonald suggested that one possible explanation for their result might derive from retrieval interference. Namely, they proposed that retroactive interference occurred at the integration site because phonological information was used during the retrieval of the displaced filler (e.g., *baker* in sentence 1a). It is not possible to evaluate this account based on the Acheson and MacDonald experiments, however, as the presence of phonological overlap throughout the sentence makes it impossible to determine whether the observed slowing occurred during *integration* of the verbs with their filler, or during the *encoding* of phonologically similar items. If interference occurs at encoding then phonology would have its central role at the level of perceptual encoding, which is consistent with evidence for the primacy of the phonetic code in storing verbal material (Liberman, Mattingly, & Turvey, 1972; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979), and for phonologically mediated lexical access (e.g., Van Orden, 1987; Lukatela & Turvey, 1994; Desroches, Newman, Joanisse, 2009). However, if phonological interference manifests during thematic integration processes, as suggested by Acheson and MacDonald, then phonology must play a direct role in routine dependency-creation procedures (e.g., retrieval). This possibility would be surprising, as the relation between the segmental phonological code and the grammar is entirely arbitrary. Moreover, extant theories of sentence processing assume that parsing is primarily grammar-driven; we are not aware of any parsing theory that assigns a decisive role to segmental phonology.[1] Consequently, finding phonological interference during thematic integration would be highly significant.

Evaluating these two possibilities depends substantially on the memory model that is assumed to support incremental sentence processing. Under theories that posit a phonologically mediated working memory (WM) store, in which incremental sentence representations are actively maintained (e.g., those that store intermediate representations in Baddeley's *phonological loop*), phonology could have a pervasive influence on stages of sentence processing beyond lexical encoding. This approach predicts that phonological interference would manifest if phonologically similar or overlapping items are held in WM. Further, such interference should increase as a function of the number of overlapping items, especially if the parser requires that entire constituents be maintained in WM during the

[1]We acknowledge that suprasegmental phonology, such as prosodic phrasing, may play an important role in parsing, particularly in the presence of phrasal ambiguities. The current paper is concerned only with word-level phonological overlap, of the sort investigated by Acheson & MacDonald (2011) and the other studies cited in the introduction.

integration of grammatical dependents (e.g., Caramazza, Berndt, & Koller, 1981; Gibson, 2000; Shankweiler & Crain, 1986).

An alternative approach assumes a sharply limited focus of attention, which constitutes active memory (e.g. Lewis, Vasishth & Van Dyke, 2006; McElree, Foraker, & Dyer, 2003). On this view, an item's initial encoding may rely on its phonological code, but it is then rapidly shuttled out of the focus of attention and into a general long-term memory (LTM), access to which is dominated by syntactic and semantic codes (cf. Bruce & Crowley, 1970; Kintsch & Buschke, 1969). Under this view, parsing does not require the simultaneous maintenance of numerous objects in a single WM store or the phonological loop, and the phonological form of displaced items is effectively inert for subsequent retrieval operations. Thus, these models predict that phonology's influence in incremental parsing should not extend past the encoding or lexical access stage. Phonologically overlapping distractors should not contribute to retrieval interference during later thematic integration because previously encountered (i.e., heard or read) items are accessed primarily via semantic and syntactic, but not phonological, cues.[2]

This paper aims to directly assess the role of phonology in retrieval and parsing by assessing whether phonological overlap can create retrieval interference. To accomplish this we used an experimental paradigm that has previously been used to demonstrate sensitivity to interference effects during incremental reading. Van Dyke and colleagues have shown that thematic integration of a filler with a verb is susceptible to interference from semantic associates (Van Dyke & McElree, 2006; Van Dyke, Johns, & Kukona, 2014; see also Gordon, Hendrick, & Levine, 2002). In a self-paced reading paradigm, participants were required to read sentences such as (2) in which a filler (e.g., *boat*) had to be interpreted as the object of a subsequent verb (e.g., either *sailed* or *fixed*). On half of the trials, participants also memorized a list of three distractor words (Load conditions), which they were asked to recall after completing the reading task and answering a subsequent comprehension question. The sentences in the NoLoad conditions were identical to those in the Load conditions, but these conditions were presented without a memory list. Interference was created in the Load conditions by manipulating the degree of overlap between the verb's affordances and the semantic properties of the distractor nouns. In Non-Interfering conditions, the semantic features of the verb uniquely resonated with the filler (*sailed* selects for *boat*). In Interfering conditions, both the target filler and the items in the memory load list were plausible objects of the verb (e.g. tables, sinks, trucks, and boats are all objects that can be *fixed*).

(2) {table-sink-truck} / {----- ----- -----}

It was the boat that the guy who lived by the sea **{sailed/fixed}** in two sunny days.

Van Dyke and McElree (2006) reported a significant slowdown at the verb only when participants were required to memorize a set of items that were plausible objects for the

[2]It is not our position that phonological form is *absent* from the stored memory trace of a lexical item. Indeed, there is much evidence that lexical items are stored as integrated representations that include orthographic, phonologic, and semantic information (e.g., Seidenberg & McClelland, 1989; Harm & Seidenberg, 1999; 2004).

critical verb. They reasoned that this was an effect of *retrieval* interference arising because semantic information from the verb was used as a cue to retrieve the filler from memory. Retrieval of the filler is required for integration with the verb because the representation of the filler is displaced from the focus of attention by the material that intervenes between the filler and the verb (e.g., McElree, 2000; McElree, Foraker & Dyer, 2003; Swinney et al., 1988; Wagers & Phillips, 2013).[3] Because the paradigm provides an index of retrieval interference, it can be used to test whether phonological overlap among potential filler noun phrases engenders a similar detrimental effect on argument integration.

The current study employs this paradigm in two experiments to test the effects of phonological overlap between a set of distractor words and a filler. As in the original study, participants read a sentence containing a filler which must be integrated with its verb and were required to maintain a list of memory words. In Experiment 1 two conditions tested the role of phonological overlap: in the overlapping condition, the memory words rhymed with the filler (*coat-vote-note*) while in the non-overlapping condition, the memory words did not rhyme (*table-sink-truck*). Experiment 2 also contained these conditions, but added a further control condition in which the memory words rhymed with each other, but not with the filler. This allowed us to more precisely assess the source of any interference effects. Based on Van Dyke and McElree's findings, interference effects should manifest at or after (but not before) the retrieval site of the filler if phonological information is used as a cue for retrieval. On the other hand, if effects arise earlier, then these must be due to pre-retrieval encoding processes, which are not related to thematic argument integration.

As in the original Van Dyke & McElree study, both experiments reported here used a region-by-region self-paced reading task to monitor the time-course of processing, This approach offers a key improvement over previous designs. It permits the onset of any interference effects to be more easily localized; many previous studies of phonological overlap, which have used offline, or full-sentence reading times as their dependent measures, and have therefore been unable to precisely pinpoint the origin of phonological interference effects (e.g., Baddeley et al., 1981; McCutchen et al., 1991). In addition, because this design incorporates distinct regions associated with the initial processing of the filler and regions associated with the integration of the filler with a verb, we are able to distinguish potential interference in encoding and retrieval stages. Previous studies of phonological overlap have been unable to cleanly determine whether such effects arise during encoding or retrieval processes (e.g., Acheson & MacDonald, 2011).

[3]An alternative view suggests that an unintegrated filler might not require retrieval at the gap site if it is maintained in active memory until its gap position is encountered. This view seems to be the dominant interpretation of the *Active Filler Strategy* (e.g., Clifton & Frazier, 1989; Fodor, 1995) which was motivated, in part, by immediate effects of the filler at a potential gap site (e.g., the filled gap effects of Stowe, 1986). We note, however, that a retrieval-driven account of filler-gap processing is not necessarily at odds with an Active Filler strategy. Such a strategy can be implemented in a variety of different architectures, including some which would require retrieval of the the un-integrated filler (see McElree, Foraker & Dyer, 2003 for discussion). Under the retrieval view, an active filler strategy minimally entails that a filler should be marked for subsequent retrieval as soon at is first encountered and that retrieval should be automatically triggered when the parser determines that a gap site is possible.

# Experiment 1

## Method

**Participants—**54 participants of college age (mean = 20.66; range = 16-24) were recruited from the New Haven community and compensated at a rate of $20 per hour. Because our sample population was drawn from the local community and therefore included a number of non-students, individual differences in performance on a number of cognitive measures were much broader than is typically observed among participants recruited through a university subject pool. To restrict the range of some inter-individual differences, participants were screened using the sight-word subtest of the Tests of Word Reading Efficiency (TOWRE; Torgesen et al., 1999), which is a standard clinical tool for assessing reading fluency and accuracy. A raw score of 90 was required for inclusion in the study. We screened participants in this manner because word-level reading difficulty is a potential confound when interpreting reading times as an index of higher-order integration processes. Moreover, word reading difficulties are often associated with phonological processing deficits (e,g., Melby-Lervåg, Lyster, & Hulme, 2012; National Reading Panel, 2000), which could also present a confound for interpreting effects of our phonological manipulation. This resulted in a sample population of 24 participants (mean age = 20.65; range = 16–24). TOWRE raw scores ranged from 92-104; mean = 100.75; standard deviation = 4.58. This corresponds to an average grade equivalency of > 12$^{th}$ grade, and an average age equivalency of 17.4 years.

### Design and Materials

**Sentences:** Test sentences were 30 object cleft constructions that contained a filler (e.g., *the boat*) that had to be retrieved and integrated with a subsequent and linearly distant verb (e.g., *sailed*). Examples are shown in Table 1. The filler was always the head noun of the cleft. The experiment had three conditions, which manipulated whether participants were required to maintain a list of words during sentence reading for later recall, as well as properties of the list items. There were two Load conditions. In the *NoRhymeLoad* condition the list was comprised of three words that had no phonological overlap with the filler. In the *RhymeLoad* condition, list words rhymed with the filler. Finally, in the *NoLoad* condition participants were not given a list of words prior to reading the sentence.

**Memory Lists:** The memory words in the RhymeLoad and NoRhymeLoad lists were nouns that were semantically unrelated to the main verbs of the experimental sentences. Words in the RhymeLoad lists were nouns that rhymed with the sentential filler. Words in the NoRhymeLoad lists were nouns that were phonologically unrelated to the filler. Both lists were matched for word length, number of syllables, and log frequency (SUBTL database; Brysbaert & New, 2009), so that no incidental differences between them would contribute to any observed effects. In addition, both lists were also matched in length, syllable count, and log frequency with their corresponding filler.

In selecting the nouns for the RhymeLoad memory list, we chose to focus on a single dimension of phonological overlap – the rime, rather than the onset, which was the focus in much previous work (e.g., Acheson & MacDonald, 2011; McCutchen et al., 1991). We adopted this approach for three reasons. First, there is evidence that the structure of phonological neighborhoods in English is dominated by rime neighbors, rather than by consonant or lead neighbors, for both single and multisyllable words (De Cara & Goswami, 2002). Second, rime neighbors tend to have greater phonological overlap than other types of neighbor (Ziegler & Goswami, 2005). Thus, by focusing on rime neighbors, we created a very high degree of phonological overlap between list items and filler (as well as intra-list overlap). Finally, there is evidence that phonological overlap in the onset of words (e.g., baker – banker) elicits inhibition related to initial lexical access (e.g., Goldinger, 1999; Monsell & Hirsh, 1998; Slowiaczek & Hamburger, 1992; Slowiaczek & Pisoni, 1986) while overlap in the rime facilitates lexical access (e.g., Burton, 1992; Slowiaczek, McQueen, Soltano, & Lynch, 2000). Hence, we believed the rime manipulation provided the best means for generating the desired phonological interference from our RhymeLoad memory list.

**Filler sentences:** In addition to the experimental items, 90 filler items, selected from those used by Van Dyke & McElree (2006), were distributed between test items in a pseudo-randomized fashion, to ensure that participants did not see two test items in a row. These were designed to discourage participants from associating either the cleft construction or the presence of a memory list with the experimental conditions. Accordingly, some of these were subject clefts, half of which also had an associated memory list containing words unrelated to the verbs in the sentences they preceded (e.g., memory list = *kettle*, *timeline*, *magnet*; Sentence = *It was the warden who worked in the state prison who discovered the escape tunnel in the prison library*.). Other filler items consisted of non-clefted right-branching structures, half of which had associated memory lists; again, these had no relationship to the verbs in the sentence (e.g., memory list = *muffin*, *basin*, *theater*; Sentence = *The bartender commented that the patron grumbled that the drunkard enjoyed the booze*.). These filler together with the experimental items totaled 120 items that were shown to each participant.

**Comprehension questions:** Participants answered a “Yes”-“No” comprehension question after every sentence. Comprehension questions probed different aspects of the previous sentence across items. Thirty percent of questions paired with test sentences directly probed the interpretation of the filler-gap dependency between the head of the cleft and the final verb (e.g. the sentence *It was the flower that the wife who purchased the new necklace planted after the heavy rainfall* was followed by the question *Did the wife plant a shrub?*). The remainder of the questions associated with the test sentences targeted other relations in the sentence (e.g., the verb phrase in the relative clause attached to the embedded subject, as in the example in Table 1). Questions on filler trials similarly queried the filler-gap dependency or some other aspect of the sentence. Correct responses to comprehension questions were divided evenly between “Yes” and “No”.

**Procedure—**Participants visited Haskins Laboratories and completed the experiment, and the TOWRE in a single session. The experiment was run on a desktop PC using the E-Prime software package (Schneider, Eschman, & Zuccolotto, 2002). In the NoLoad condition, each trial began with the self-paced reading display. Participants read each sentence using a self-paced phrase-by-phrase display (phrase boundaries are shown in Table 1), advancing through each sentence by pressing the space bar. A comprehension question followed the end of the test sentence. Participants could respond Yes or No to the comprehension questions by pressing the “1” or “3” keys on the number pad. After answering the comprehension question, the next trial began. In Load conditions, participants first viewed a screen containing the three memory words. Load words were displayed simultaneously on a single line and separated from each other by five dash characters “-----”. Presentation of the load list terminated automatically after 5000ms. After terminating the load presentation, participants read the sentences and answered the comprehension questions as in the NoLoad condition. After answering the comprehension question, participants were asked to recall the memory words from the list by typing them into an answer field in the same serial order in which they were presented. Participants could skip a position if they did not remember a word. In addition, particpants were instructed that they should be as accurate as possible on both the memory recall task and the sentence comprehension task. The next trial began immediately after entry of the last memory word (or entry of the enter key if they did not recall the word).

**Analysis—**Analyses were conducted using linear and logistic mixed effects models. The factor *Condition* was Helmert coded in order to enable two contrasts to be tested: one for an overall effect of Condition (contrasts: -1,1,0 for RhymeLoad, NoRhymeLoad, and NoLoad, respectively) and the second for an effect of maintaining a Load list (-1,-1,2). Models with maximal random effects structures were used, providing random slopes and intercepts for subject and item (Barr, Levy, Scheepers and Tily, 2013). In this and all subsequent experiments, we report the *t*-values or *z*-values associated with our analyses. We adopt the convention whereby any effect whose absolute *t*-value exceeds 2 is considered significant (Gelman & Hill, 2007).

Reading times that exceeded a threshold of 3500ms were discarded, resulting in the loss of less than 1.5% of the total data. Of the remaining data, reaction times that exceeded a threshold of 2.5 standard deviations from the mean by region and condition were discarded (Ratcliff, 1993).

## Results

### Comprehension Question and Recall Accuracy

There was no effect of Load on participants' comprehension question accuracy (logistic mixed effects model, $z < 2$). Comprehension was numerically higher, on average, in the NoLoad condition (90% of all questions were answered correctly) than in either the RhymeLoad or NoRhymeLoad conditions (average accuracies of 86% and 85%, respectively).

We assessed recall accuracy using a strict criterion, such that recall on a trial was considered accurate only when participants reproduced all of the words in the load list in the correct order. Between the Load conditions, there was no reliable difference in the accuracy with which participants were able to recall items in the load list (71% in the RhymeLoad condition, 70% in the NoRhymeLoad condition). Recall accuracy was also not significantly different under a more liberal criterion that did not enforce accurate order of the memoranda.

### Reading Time Results

Table 2 presents region-by-region reading times. Table 3 and presents a summary of statistical effects by region.

### Region 1 - sentence onset

At the sentence-initial position there was no significant effect of Condition or Load.

### Region 2 - post filler

We observed a significant effect of Condition ($t$ = -2.73) in the region immediately following the filler. Pairwise comparisons revealed that mean RTs were longer in the RhymeLoad condition than in the the NoRhymeLoad condition ($t$ = -2.96) and the NoLoad condition ($t$ = -2.34). RTs in the NoRhymeLoad and NoLoad conditions did not differ significantly from one another. Thus, this effect is attributable to disproportionately long reading times in the RhymeLoad condition relative to the other two conditions.

### Region 3 - relative pronoun

Reading times did not differ significantly by condition; however, average reading times in Load conditions were marginally slower than those in the NoLoad condition ($t$ = -1.86).

### Region 4 - RC verb

Reading times at the embedded RC-verb did not differ significantly as a function of Condition or Load.

### Region 5 – RC spillover

There were no significant effects following the RC-internal verb.

### Region 6 – critical verb

Average reading times did not differ across conditions at the retrieval verb.

### Region 7 – sentence final spillover

At the sentence-final spillover region Load conditions were read more quickly than the NoLoad condition ($t$ = 2.49). Reading times did not differ significantly between the two Load conditions.

## Discussion

Experiment 1 tested whether retrieval of a filler was subject to interference from sentence-external phonological competitors in a memory load list. The study manipulated (i) whether participants were required to maintain a list of distractors and (ii) whether the phonological features of the distractor items overlapped with the filler. Our analyses showed that mean reading times were longer in the region immediately following the filler (Region 2) when participants were required to maintain a list of rhyming distractors. Reading times at the critical verb did not differ significantly across conditions as a function of Load or Load type In the sentence-final spillover region (Region 7), reading times were shorter in both conditions where participants were required to maintain a Load list. Importantly, this effect of Load was not modulated by phonological overlap. A similar effect was reported in Van Dyke & McElree (2006) and Van Dyke, Johns & Kukona (2014), both of which also used this paradigm. As in those studies, we suggest that this effect is likely due to participants rushing to get to the recall portion of the trial.

The crucial region for assessing retrieval interference was at the critical verb (Region 6). For phonological overlap to affect subsequent integration of the to-be-retrieved filler, a difference between the RhymeLoad condition and all others should have emerged here, or perhaps later. However, there was no significant effect of phonological overlap in these regions, lending no support to the proposal that phonological cues are used during thematic integration, or that phonological overlap produces retrieval interference.

Although no phonological interference effects were observed at the point of filler retrieval, there was a reliable effect of phonological overlap earlier in the sentence. In two regions following the introduction of the filler (Regions 2 and 3), reading times were longest when participants had to maintain a list of distractors that rhymed with the filler. Reading times did not differ between NoRhymeLoad and NoLoad conditions in the same region, indicating that increased processing times in this region should not be attributed to a general effect of maintaining distractor words. Because the elevated reading times do not coincide with the point at which the filler must be retrieved, they cannot be construed as retrieval interference, or as arising from use of phonological retrieval cues during integration. The time course of the effects suggests instead that phonological interference effects emerge as a proactive encoding effect: encoding a representation of the filler NP is more difficult when rhyming words are active in recent memory.

## Experiment 2

We have suggested that Experiment 1 indicates that the locus of phonological interference effects is at *encoding*: the presence of memory words that rhyme with the filler create additional difficulty for storing the lexical entry of the filler. However, our results thus far are also consistent with an alternative account, in which the observed slowdown may be due simply to greater difficulty associated with a list of rhyming items. Certainly, there is plenty of evidence in the memory literature, using a variety of tasks, for the so-called “phonological similarity effect”, by which lists containing a high degree of phonological overlap are not recalled with the same facility as lists with phonologically dissimilar items (e.g., Baddeley, 1966; Craik, 1968; Conrad & Hull, 1964; LaPointe & Engle, 1990; Surprenant, Neath, & LeCompte, 1999; Wickelgren, 1965). In addition, there is evidence that lists with a length of three items, as employed in our first experiment, are the minimum length for memory performance to be affected (Baddeley, Lewis, & Vallar, 1984). Thus, the goal of this experiment is to tease apart these two possibilities by assessing the impact of maintaining a rhyming list that does and does not overlap with the filler. As in the previous experiment, we again used lists that did and did not contain rhyming words. However, unlike the previous experiment, we orthogonally varied the form of the filler, such that it either did or did not rhyme with the lists.

If the post-filler difficulty observed in the previous experiment can be attributed to the cost of maintaining a rhyming list, rather than interference between list contents and the sentential filler, we should observe a main effect of the type of memory list: that is, when the memory list is comprised of rhyming words, readers' post-filler RTs should be slower regardless of whether the filler's phonological characteristics overlapped with the memory list. Under this interpretation, both form of the filler and its relationship to the memory list words is immaterial to the presence of the early processing difficulty; rather, the observed difficulty would result from the list-internal similarity of the load words themselves. By contrast, if the difficulty is linked to encoding the filler under conditions of phonological interference, we would expect slowdown to be restricted to the post-filler region only when the filler had phonological overlap with the memory list.

## Method

### Participants

39 college students (mean age = 19.8, range = 18-24) participated in the study. Participants were compensated at a rate of $20 per hour. We restricted our sample population to college-going individuals to enhance the comparability of our results with previous research. As in Experiment 1, in order to exclude participants with word reading difficulty we analyzed data only for those participants whose raw score on the sight-word subtest of the TOWRE was greater than 90 (mean score = 101.9, range = 92-104). This resulted in the exclusion of 7 participants, leaving a total of 32 in our current sample.

### Design and Materials

**Sentences—**32 test items were created, 27 of which were adapted from the previous experimental items. Test sentences were identical in structure to those in the previous experiments. Test items followed a 2 × 2 factorial design that manipulated the lexical content of the initial filler (OldFill *v.* NewFill) and whether words in a memory load list rhymed with each other (RhymeLoad v. NoRhymeLoad). Thus, unlike the previous experiments, all test items were preceded by a memory list. In OldFill conditions, the test sentence was identical to the test sentence in the previous experiments. NewFill conditions were created by exchanging the filler noun in the original sentence with a semantically similar noun of comparable length and frequency, but which did not rhyme with any of the words in the memory lists (e.g. *boat* in Table 1 was changed to *ship*). These two were crossed with a load list in which all of the memory words either rhymed with one another (RhymeLoad) or did not (NoRhymeLoad). An example test item is provided in Table 4.

**Memory lists—**The memory lists for the 27 items taken from the previous experiments were unchanged. The lists for the remaining 5 items were created according to the same parameters as the previous experiments, i.e., rime manipulation and controlling for length, number of syllables, and log frequency. (Thus, for example, the condition in which RhymeLoad and OldFill are crossed is equivalent to the RhymeLoad condition from the previous experiments.)

**Filler sentences—**Test sentences were pseudo-randomly distributed among 88 filler sentences such that participants read 120 items in Experiment 2, just as in Experiment 1. 80 fillers were taken directly from the previous experiment and 8 were created for Experiment 2 according to the same guidelines as previous fillers.

**Comprehension questions—**Experiment 1 queried the long-distance dependency 30% of the time. In addition, we designed this experiment to more rigorously probe comprehension of the critical filler-gap relation. Comprehension questions directly probed the interpretation of the filler-gap dependency on 50% of trials.[4] When possible, a rhyming distractor word was substituted for the filler in the comprehension question (e.g. the sentence *It was the cake that the secretary who commuted by bus baked before the office party* was followed by the question *Did the secretary bake a steak?*, because *steak* was one of the distractor words). As in the previous experiments, the remainder of the questions addressed other non-critical sentential relations.

### Procedure

The procedure was identical to that used for Experiment 1.

### Analysis

As in the previous experiments, all analyses were conducted using mixed effects models with maximal random effects structures unless otherwise noted. The factors List Type (RhymeLoad v. NoRhymeLoad) and Filler Type (OldFill v. NewFill) were centered and sum-coded (range -0.5, 0.5) in order to minimize collinearity and so that reported regression coefficients could be interpreted similarly to ANOVA main effects and interactions.

## Results

### Comprehension Question and Recall Accuracy

There were no significant effects or interactions in comprehension accuracy; participants answered comprehension questions correctly on 86% of trials. As in the previous experiment, there was no reliable difference in strict recall accuracy between conditions (average strict recall accuracy: 73%), and average recall accuracy again did not differ with a more lenient criterion.

### Reading Time Results

Tables 5 and 6 present average reading times by region and summaries of statistical effects, respectively.

### Region 1 - sentence onset

In the first region, average reading times in the RhymeLoad conditions were higher than in the NoRhymeLoad condition ($t = 2.479$). There was no effect of Filler Type, and no interaction.

[4]We thank Jane Ashby (p.c.) for requesting this change, which bolsters the reliability of our results and interpretations.

### Region 2 - post filler

Just after the filler, reading times were again longer on average in the RhymeLoad conditions ($t$ = 2.798). Unlike the effect of list type in the previous region, this main effect was qualified by a significant List Type × Filler Type interaction ($t$ = 2.520). The average reading time in the RhymeLoad-OldFill condition was significantly longer than in the RhymeLoad-NewFill condition (pairwise $t$ = 2.743), the NoRhymeLoad-OldFill ($t$ = 3.453), and NoRhymeLoad-NewFill condition ($t$ = 3.123). All other pairwise comparisons were not significant (all $t$s < 1).

### Regions 3 – 5

There were no significant effects or interactions in the relative pronoun, RC verb, or RC spillover regions.

### Region 6 – critical verb

At the critical verb, average reading times did not differ across conditions.

### Region 7 – sentence final spillover

RhymeLoad conditions were read more quickly than NoRhymeLoad conditions in the final region of the sentences ($t$ = -2.501). Numerically, this effect was carried by the RhymeLoad-OldFill condition (see Table 4), but the interaction of Load Type and List Type was not significant.

## Discussion

Experiment 2 was designed to test whether phonological interference effects that followed the filler in Experiment 1 should be attributed to the general cost of maintaining a rhyming list or whether the effects should be specifically linked to the overlap between memory list words and the filler. Test sentences were preceded by a memory list whose members rhymed with one another or with words that did not rhyme. The lexical content of the filler was also manipulated, such that it would rhyme with memoranda in one rhyme list, but not in another.

Participants in Experiment 2 read the onset of the sentence more slowly when maintaining a rhyming list. However, participants spent more time reading the region immediately following the filler only when the memory words and the filler rhymed; there was no comparable slowdown when the filler did not rhyme with the memoranda. This effect of phonological interference parallels the effect of phonological interference in Experiment 1. This finding supports our assertion that the phonological interference effect in Experiment 1 is conditioned on the interaction of the content of the memory list with the filler. Further, the finding that the phonological overlap had no consequences for recall performance, but had early online effects during reading, also provides support for our characterization of this as an effect of proactive interference of the rhyming list words on *encoding*.

The findings of Experiment 2 also corroborate the findings of Experiment 1 in that they offer no support for the hypothesis that phonologically similar distractors interfere with retrieval for dependency creation. As in the previous experiments, phonological overlap did not have any discernible impact on the resolution of the filler-gap dependency as measured in reading times at the verb or spillover region.

## General Discussion

The goal of this study was to investigate the potential role phonological information may play during incremental sentence processing. In particular, we asked whether phonological cues are used in the retrieval of previously processed constituents for the purposes of thematic integration. Our method for examining this issue was to assess the presence of interference from phonologically similar distractors held in memory during online sentence processing, using a paradigm that has previously been used to demonstrate retrieval interference effects.

Our results provide no evidence that phonological information interferes with retrieval during the creation of long-distance argument dependencies. Phonological overlap between a retrieval target and distractor words in a memory load list did not increase the difficulty of retrieving or integrating the target. In this regard, phonological information contrasts with syntactic and semantic cues, which have been shown to engender retrieval interference in the construction of long distance dependencies (Van Dyke, 2007; Van Dyke, Johns, & Kukona, 2014; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006; 2011). Our findings are consistent with a model of the parser that exclusively uses morphosyntactic and semantic cues to guide memory access during structure building and integration (e.g. Lewis & Vasishth, 2005; Lewis, Vasishth & Van Dyke, 2006; Van Dyke & Lewis, 2003).

Although we found no effects of phonological overlap on the creation of grammatical dependencies, we did observe consistent interference effects early in our sentences in both experiments. We observe these effects only when there was phonological overlap between the distractors in memory and the sentential filler. It is possible, in principle, that the observed effects reflect increased difficulty in maintaining the rhyming memoranda; that is, that the phonologically overlapping filler elicited a phonological similarity effect, interfering with list memory maintenance. However, in all experiments recall performance did not vary across conditions, suggesting that the observed effects reflect the rhyming lists' interference with online sentence processing, and not vice versa.

Since the time-course of the phonological interference effects in these experiments is clearly inconsistent with retrieval and/or integration difficulty, they most likely reflect difficulty with either lexical access or encoding. Regarding lexical access, it is uncontroversial that the initial stages are at least partially mediated through a phonological code (Van Orden, 1987; Perfetti & Bell, 1991; Lukatela & Turvey 1994; Harm & Seidenberg 2004; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Desroches, Newman & Joanisse 2009). Hence, the idea that the phonological characteristics of the memory words interfered with lexical access may appear to be attractive. However, numerous studies have shown that phonological overlap in the rime *facilitates* lexical access (e.g., Fallon, Groves, & Tehan, 1999; Frankish, 1995; Slowiaczek et al., 2000; see Rapp & Samuel, 2002, for a useful discussion of the use of rime or coda information in lexical access). Given our exclusive manipulation of the phonological overlap of the rime, then, it is likely that the effect of our phonological manipulations (if any) on lexical access should have been facilitative – a pattern that is plainly absent in both of the experiments. Consequently, we believe the more likely explanation is that the phonological interference effects in these experiments reflect proactive interference on encoding, such that the contents of the rhyming memory lists rendered the encoding processes associated with the phonologically similar filler during online processing more effortful.

Our finding of reliable phonological overlap effects is in part consistent with previous work observing similar effects. However, our explanation of these effects as due to encoding interference is a departure from previous work that has focused on phonology's role in maintaining order information as the source of these effects (e.g., Shankweiler et al. 1979; Shankweiler & Crain 1986; Acheson & MacDonald, 2011). Indeed, findings from list-learning experiments in the memory literature have shown that phonological similarity in a memory list affects memory for the *order* of items in the list and not memory for the items themselves (e.g., Wickelgren, 1965; Fallon et al., 1999). Thus, if order information is important during parsing—for example as part of a partially processed constituent that must be actively maintained in working memory—then one would expect to see phonological overlap effects in comprehension.

In contrast, our results show that comprehension, as measured by the ability to integrate dependent constituents, does not require phonological information. This is consistent with the prediction of cue-based parsers, which do not represent or store linear order information explicitly (Lewis, 2000; Lewis & Vasishth, 2005; Lewis, Vasishth & Van Dyke 2006; Van Dyke & Lewis, 2003); thus, according to this class of parser, order information should not be able to guide parsing and retrieval. For example, items cannot be retrieved based on their linear position relative to another item (McElree, Foraker & Dyer, 2003). Motivation for this position also comes from findings from the memory literature that demonstrate the slow, serial nature of accessing order information (McElree & Dosher, 1993; McElree, 2006), making it incompatible with the immediate, incremental nature of real-time sentence parsing. Instead, the parser relies on its ability to predict syntactic structure to generate distinctions among elements. To illustrate, in the high overlap sentence (1a; repeated below) investigated by Acheson & MacDonald (2011), the NP *baker* can be distinguished from the NP *banker* by the verb expectation associated with it: *baker* is predicted to be the head of a yet-to-occur matrix verb, while the structure surrounding *banker* suggests it will be the head of a yet-to-occur embedded verb with a missing argument. (See Lewis & Vasishth (2005), for computational implementation of this idea.)

(1a) The baker that the banker sought bought the house.

Thus, although phonologically based encoding interference may occur while processing *banker* – and indeed, the current results suggest that it should – the phonological overlap should not in itself affect the proper assignment of grammatical dependencies (e.g., retrieval and integration of *banker* as the object of *sought*). We note in passing that this position is also consistent with findings in the aphasia literature, where patients with phonological deficits can nevertheless process syntactically complex sentences (Butterworth, Campbell, & Howard, 1986; Saffran & Martin, 1990; Waters, Caplan, & Hildebrandt, 1991).

Although neither serial order not phonological cues are critical elements for a cue-based parser in the typical case, we nonetheless hypothesize that phonological representations may play a role in situations of high syntactic and semantic interference (i.e., when the parser encounters difficulty adequately distinguishing between two items in memory due to a high degree of syntactic and semantic cue-overlap). In this case, the parser shifts into a deliberate "repair mode", during which the parser might "check its work" by retrieving previously processed items and using them to re-parse the sentence. Phonological codes – as an integral part of the representations of these retrieved items - could be employed at some step in the repair process. One possibility is that the parser exploits phonological information to reconstitute order information for the purposes of the reparse. Greater phonological overlap between constituents would result in a less veridical recreation of the original sentence order. We speculate that this might occur precisely in cases of reversible relative clauses, such as the overlap and non-overlap conditions in Acheson and MacDonald (2011). Here, the transposition of the order of the two NPs *the baker* and *the banker* would result in a coherent, though erroneous, interpretation. The uncertainty created by the high overlap, together with the lack of constraining semantic evidence, may be sufficient to trigger a deliberate review in order to inspect the relative positioning of the two NPs (which could, in turn, provide confirmatory evidence in support of supposed syntactic positions). Yet, as noted above, phonological similarity affects memory for the order of items in the list (and not for the items themselves). Hence, the phonological overlap will promote the creation of an inaccurate reconstruction, which may lead to elevated reading times and poorer comprehension – the very pattern observed by Acheson and MacDonald.

This hypothesis also allows us to reconcile the fact that phonological overlap did not result in comprehension deficits in our experiments, while other studies did observe this effect. We believe that the degree of syntactic and semantic interference in our materials was never high enough to require consulting a reconstructed phonological representation. That is, our sentence materials could be processed using automatic processing routines without triggering repair or reanalysis routines. Thus, phonological overlap in and of itself is not sufficient to produce poor comprehension.

To conclude, our findings suggest that phonological information does not play a role in the retrieval of previously processed items for dependency creation during routine parsing and sentence comprehension. More narrowly, our findings argue against explaining phonological overlap effects as reflecting retrieval interference. Instead, we consider phonological overlap effects to arise from two sources. First, phonological overlap between previously processed items and items currently in the focus of attention can lead to encoding interference, as was reported here. Secondly, overlap effects such as those reported previously in the literature (e.g., Acheson & MacDonald, 2011; McCutcheon et al., 1999) may arise when neither syntactic nor semantic cues are sufficient to distinguish similar constituents. In this case, we speculate that the parser may shift from automatic processing into a conscious repair mode, and evoke a phonologically mediated representation of the sentence's linear sequence in order to verify a supposed interpretation. Thus, while phonological interference may play a role in comprehension when routine processing becomes onerous, it remains outside the domain of typical argument dependency formation.

## Acknowledgments

This work was supported by the following grants from the NIH National Institute of Child Health and Human Development to Haskins Laboratories: HD 073288 (Julie A. Van Dyke, PI), HD 001994 (Jay Rueckl, PI), and HD 067364 (Kenneth R. Pugh, PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors are grateful to Morgan L. Bontrager, Hannah R. Jones, and Eva Montas for assistance with data collection and materials development.

## References

Acheson DJ, MacDonald MC. The rhymes that the reader perused confused the meaning: Phonological effects during on-line sentence comprehension. Journal of Memory and Language. 2011; 65(2): 193–207. [PubMed: 21743771]

Baddeley AD. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. Quarterly Journal of Experimental Psychology. 1966; 18:362–365. [PubMed: 5956080]

Baddeley AD, Eldridge M, Lewis V. The role of subvocalisation in reading. Quarterly Journal of Experimental Psychology A: Human Experimental Psychology. 1981; 33:439–454.

Baddeley, AD.; Hitch, G. Working memory. In: Bower, GH., editor. The psychology of learning and motivation: Advances in research and theory. New York, NY: Academic Press; 1974. p. 48-79.

Baddeley A, Lewis V, Vallar G. Exploring the articulatory loop. The Quarterly Journal of Experimental Psychology. 1984; 36(2):233–252.

Baddeley, AD.; Vallar, G.; Wilson, B. Sentence comprehension and phonological memory: Some neuropsychological evidence. In: Coltheart, M., editor. Attention and performance: The psychology of reading. Vol. XII. Hillsdale, NJ: Lawrence Erlbaum Associates; 1987. p. 509-529.

Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language. 68(3):255–278.

Bruce D, Crowley JJ. Acoustic similarity effects on retrieval from secondary memory. Journal of Verbal Learning and Verbal Behavior. 1970; 9(2):190–196.

Brysbaert M, New B. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods. 2009; 41(4):977–990. [PubMed: 19897807]

Butterworth B, Campbell R, Howard D. The uses of short-term memory: A case study. Quarterly Journal of Experimental Psychology. 1986; 38(4):705–737. [PubMed: 3809577]

Caramazza A, Berndt RS, Koller JJ. Syntactic processing deficits in aphasia. Cortex. 1981; 17(3):333–347. [PubMed: 7333108]

Conrad R, Hull AJ. Information, acoustic confusion, and memory span. British Journal of Psychology. 1964; 55:429–432. [PubMed: 14237884]

Craik FI. Two components in free recall. Journal of Verbal Learning and Verbal Behavior. 1968; 7:996–1004.

De Cara B, Goswami U. Similarity relations among spoken words: The special status of rimes in English. Behavior Research Methods, Instruments, & Computers. 2002; 34(3):416–423.

Desroches AS, Newman RL, Joanisse MF. Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. Journal of Cognitive Neuroscience. 2009; 21(10):1893–1906. [PubMed: 18855555]

Fallon AB, Groves K, Tehan G. Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT, but not MAN. International Journal of Psychology. 1999; 34:301–307.

Frankish C. Intonation and auditory grouping in immediate serial recall. Applied Cognitive Psychology. 1995; 9:5–22.

Gelman, A.; Hill, J. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press; 2006.

Gibson, E. The dependency locality theory: A distance-based theory of linguistic complexity. In: Marantz, A.; Miyashita, Y.; O'Neil, W., editors. Image, Language, Brain. Cambridge, MA: MIT Press; 2000. p. 95-126.

Goldinger SD. Only the shadower knows: Comment on Hamburger and Slowiaczek (1999). Psychonomic Bulletin & Review. 1999; 6:347–351. [PubMed: 12199221]

Gordon PC, Hendrick R, Levine WH. Memory-load interference in syntactic processing. Psychological Science. 2002; 13:425–430. [PubMed: 12219808]

Haber LR, Haber RN. Does silent reading involve articulation? Evidence from tongue twisters. American Journal of Psychology. 1982; 95:409–419. [PubMed: 7180948]

Hannagan T, Magnuson J, Grainger J. Spoken word recognition without a TRACE. Frontiers in Psychology. 2013; 2(4):1–17.

Harm MW, Seidenberg MS. Reading acquisition, phonology, and dyslexia: Insights from a connectionist model. Psychological Review. 1999; 106:491–528. [PubMed: 10467896]

Harm MW, Seidenberg MS. Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. Psychological Review. 2004; 111(3):662–720. [PubMed: 15250780]

Keller TA, Carpenter PA, Just MA. Brain imaging of tongue- twister sentence comprehension: Twisting the tongue and the brain. Brain and Language. 2003; 84:189–203. [PubMed: 12590911]

Kennison SM, Sieck JP, Briesch KA. Evidence for a late-occurring effect of phoneme repetition during silent reading. Journal of Psycholinguistic Research. 2003; 32(3):297–312. [PubMed: 12845941]

Kintsch W, Buschke H. Homophones and synonyms in short-term memory. Journal of Experimental Psychology. 1969; 80(3):403–407.

LaPointe LB, Engle RW. Simple and complex word spans as measures of working memory capacity. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1990; 16:1118–1133.

Lewis RL, Vasishth S, Van Dyke JA. Computational principles of working memory in sentence comprehension. Trends in Cognitive Sciences. 2006; 10(10):447–454. [PubMed: 16949330]

Lewis, RL. Specifying architectures for language processing: Process, control, and memory in parsing and interpretation. In: Crocker, MW.; Pickering, M.; Clifton, C., Jr, editors. Architectures and Mechanisms for Language Processing. Cambridge, UK: Cambridge University Press; 2000. p. 56-89.

Lewis RL, Vasishth S. An activation-based model of sentence processing as skilled memory retrieval. Cognitive science. 2005; 29(3):375–419. [PubMed: 21702779]

Lukatela G, Turvey MT. Visual lexical access is initially phonological II: Evidence from phonological priming by homophones and pseudohomophones. Journal of Experimental Psychology: General. 1994; 123(4):331–353. [PubMed: 7996120]

Magnuson JS, Dixon JA, Tanenhaus MK, Aslin RN. The dynamics of lexical competition during spoken word recognition. Cognitive Science. 2007; 31(1):133–156. [PubMed: 21635290]

Marslen-Wilson W, Zwitserlood P. Accessing spoken words: The importance of word onsets. Journal of Experimental Psychology: Human perception and performance. 1989; 15(3):576–585.

Martin N, Saffran EM. Repetition and verbal STM in transcortical sensory aphasia: A case study. Brain and Language. 1990; 39(2):254–288. [PubMed: 2224495]

McCutchen D, Bell LC, France IM, Perfetti CA. Phoneme- specific interference in reading: The tongue-twister effect revisited. Reading Research Quarterly. 1991; 26:87–103.

McCutchen D, Dibble E, Blount MM. Phonemic effects in reading comprehension and text memory. Applied Cognitive Psychology. 1994; 8(6):597–611.

McCutchen D, Perfetti CA. Coherence and connectedness in the development of discourse production. Text. 1982; 2:113–139.

McElree, B. Accessing recent events. In: Ross, BH., editor. The psychology of learning and motivation. Vol. 46. San Diego, CA: Academic Press; 2006. p. 155-200.

McElree B, Dosher BA. Serial retrieval processes in the recovery of order information. Journal of Experimental Psychology: General. 1993; 122(3):291–315.

McElree B, Foraker S, Dyer L. Memory structures that subserve sentence comprehension. Journal of Memory and Language. 2003; 48(1):67–91.

Melby-Lervåg M, Lyster SH, Hulme C. Phonological skills and their role in learning to read: A meta-analytic review. Psychological Bulletin. 2012; 138:322–352. [PubMed: 22250824]

Monsell S, Hirsh KW. Competitor priming in spoken word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1998; 24(6):1495–1520.

National Reading Panel. Report of the National Reading Panel: Reports of the Subgroups. Washington, DC: National Institute of Child Health and Human Development Clearing House; 2000.

Perfetti CA, Bell L. Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. Journal of Memory and Language. 1991; 30(4):473–485.

Ratcliff R. Methods for dealing with reaction time outliers. Psychological Bulletin. 1993; 114:510–532. [PubMed: 8272468]

Robinson DH, Katayama AD. At-lexical, articulatory interference in silent reading: The “upstream” tongue-twister effect. Memory & Cognition. 1997; 25(5):661–665. [PubMed: 9337584]

Saffran, EM.; Martin, N. Short-term memory impairment and sentence processing: a case study. In: Vallar, G.; Shallice, T., editors. Neuropsychological impairments of short-term memory. Cambridge, UK: Cambridge University Press; 1990. p. 428-467.

Schneider, W.; Eschman, A.; Zuccolotto, A. E-Prime: User's guide. Psychology Software Incorporated; 2002.

Seidenberg MS, McClelland JL. A distributed, developmental model of word recognition and naming. Psychological review. 1989; 96(4):523–568. [PubMed: 2798649]

Shankweiler D, Crain S. Language mechanisms and reading disorder: A modular approach. Cognition. 1986; 24(1):139–168. [PubMed: 3791919]

Shankweiler D, Liberman IY, Mark LS, Fowler CA, Fischer FW. The speech code and learning to read. Journal of Experimental Psychology: Human Learning and Memory. 1979; 5:531–545.

Slowiaczek LM, Hamburger MB. Prelexical facilitation and lexical interference in auditory word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1992; 18:1239–1250.

Slowiaczek LM, Pisoni DB. Effects of phonological similarity on priming in auditory lexical decision. Memory & Cognition. 1986; 14:230–236. [PubMed: 3736396]

Slowiaczek LM, McQueen JM, Soltano EG, Lynch M. Phonological representations in prelexical speech processing: evidence from form-based priming. Journal of Memory and Language. 2000; 43:530–560.

Surprenant AM, Neath I, LeCompte DC. Irrelevant speech, phonological similatiry, and presentation modality. Memory. 1999; 7:405–420.

Torgesen, JK.; Wagner, RK.; Rashotte, CA. TOWRE: Test of Word Reading Efficiency. Austin, TX: Pro-Ed; 1999.

Van Dyke JA, Lewis RL. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. Journal of Memory and Language. 2003; 49(3):285–316.

Van Dyke JA, McElree B. Cue-dependent interference in comprehension. Journal of Memory and Language. 2011; 65(3):247–263. [PubMed: 21927535]

Van Dyke JA, Johns CL, Kukona A. Low working memory capacity is only spuriously related to poor reading comprehension. Cognition. 2014; 131(3):373–403. [PubMed: 24657820]

Van Orden GC. A ROWS is a ROSE: Spelling, sound, and reading. Memory & Cognition. 1987; 15(3):181–198. [PubMed: 3600258]

Waters G, Caplan D, Hildebrandt N. On the structure of verbal short-term memory and its functional role in sentence comprehension: Evidence from neuropsychology. Cognitive Neuropsychology. 1991; 8(2):81–126.

Wickelgren WA. Acoustic similarity and retroactive interference in short-term memory. Journal of Verbal Learning and Verbal Behavior. 1965; 4(1):53–61.

Zhang S, Perfetti CA. The tongue-twister effect in reading Chinese. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1993; 19(5):1082–1093.

Ziegler J, Goswami U. Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. Psychological Bulletin. 2005; 131(1):3–29. [PubMed: 15631549]

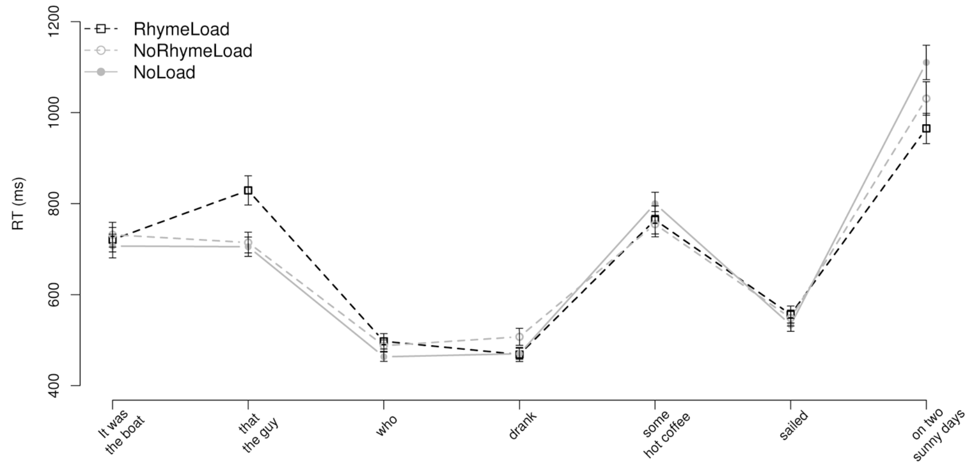


**Figure 1.**

Self-paced reading times by region for sentences in Experiment 1, in which participants held a phonologically-overlapping or non-overlapping word list, or no list at all. Error bars represent standard error of the mean reading time for each region by condition.

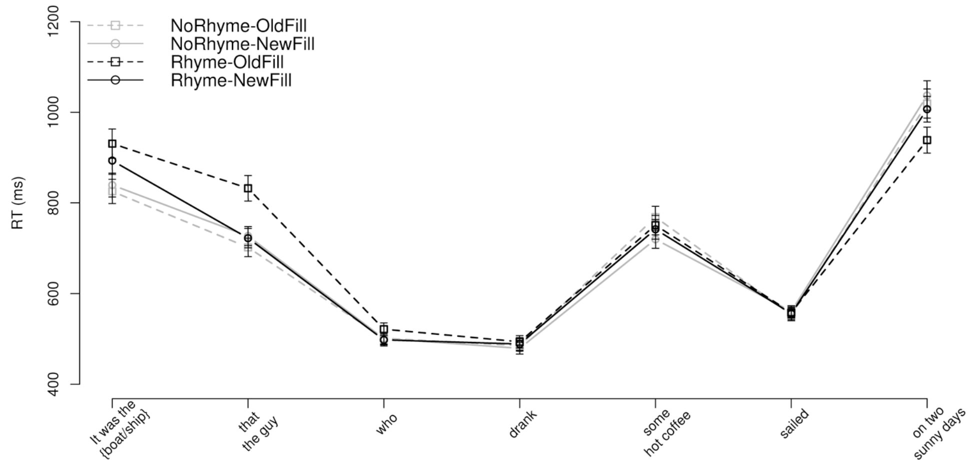


**Figure 2.**

Self-paced reading times by region for sentences in Experiment 2. Error bars represent standard error of the mean reading time for each region by condition.

Author Manuscript

Author Manuscript

Author Manuscript

## Table 1

Example Experimental item from Experiments 1 and 2. Slashes denote regions of presentation.

| Condition | Load List | Sentence | Question |
|---|---|---|---|
| NoLoad | --- --- --- | It was the boat/ that the guy/ who/ drank/ some hot coffee/ sailed/on two sunny days. | Did the guy drink hot coffee? |
| NoRhymeLoad | table-sink-truck | | |
| RhymeLoad | coat-vote-note | | |

Author Manuscript
Author Manuscript
Author Manuscript

**Table 2**

Average region-by-region reading times for Experiment 1. Standard error of the mean in parentheses.

| | Reg. 1 | Reg. 2 | Reg. 3 | Reg. 4 | Reg. 5 | Reg. 6 | Reg. 7 |
|---|---|---|---|---|---|---|---|
| RhymeLoad | 721 (27) | 829 (32) | 497 (17) | 468 (15) | 764 (31) | 556 (19) | 965 (33) |
| NoRhymeLoad | 732 (27) | 714 (23) | 488 (13) | 507 (19) | 755 (28) | 549 (17) | 1031 (38) |
| NoLoad | 707 (26) | 705 (21) | 463 (10) | 470 (12) | 800 (25) | 533 (14) | 1110 (34) |

**Table 3**

Summary of statistical effects by region for Experiment 1. Asterisks indicate significant effects. Absolute values of $t > 2$ were considered significant at a $p < .05$ level. All models contained random slopes for all fixed effects and their interaction by subject and item except for Region 4, which had only by-subject random slopes for these effects.

| | | Estimate | S.E. | t value |
|---|---|---|---|---|
| Region 1 | **(Intercept)** | 726.41 | 55.55 | 13.10* |
| | **Condition** | 3.06 | 16.67 | -0.18 |
| | **Load** | -6.48 | 10.60 | -0.61 |
| Region 2 | **(Intercept)** | 756.26 | 40.73 | 18.57* |
| | **Condition** | -57.81 | 21.17 | -2.73* |
| | **Load** | -24.17 | 14.77 | -1.64 |
| Region 3 | **(Intercept)** | 485.29 | 19.52 | 24.86* |
| | **Condition** | -7.01 | 10.59 | -0.66 |
| | **Load** | -10.48 | 5.64 | -1.86 |
| Region 4 | **(Intercept)** | 491.50 | 23.32 | 21.60* |
| | **Condition** | 12.83 | 9.05 | 1.42 |
| | **Load** | -4.76 | 5.40 | -0.88 |
| Region 5 | **(Intercept)** | 780.52 | 50.31 | 15.51* |
| | **Condition** | -1.04 | 16.63 | 0.06 |
| | **Load** | 12.10 | 10.35 | 1.17 |
| Region 6 | **(Intercept)** | 549.28 | 27.97 | 19.37* |
| | **Condition** | -4.11 | 12.84 | -0.32 |
| | **Load** | 12.10 | 7.99 | -0.86 |
| Region 7 | **(Intercept)** | 1077.1 | 68.97 | 14.62* |
| | **Condition** | 29.81 | 21.08 | 1.41 |
| | **Load** | 38.40 | 15.42 | 2.49* |

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4**

Example item from Experiment 2. Slashes denote regions of presentation.

| **Condition** | **Load List** | **Sentence** |
|---|---|---|
| NoRhymeLoad-OldFill | table-sink-truck | It was the boat /that/ the guy/ who/ lived/ by the sea/ sailed/ for two sunny days. |
| RhymeLoad-OldFill | coat-vote-note | |
| NoRhymeLoad-NewFill | table-sink-truck | It was the ship /that/ the guy/ who/ lived/ by the sea/ sailed/ for two sunny days. |
| RhymeLoad-OldFill | coat-vote-note | |

**Table 5**

Average region-by-region reading times for Experiment 2. Standard error of the mean in parentheses.

| | Reg. 1 | Reg. 2 | Reg. 3 | Reg. 4 | Reg. 5 | Reg. 6 | Reg. 7 |
|---|---|---|---|---|---|---|---|
| Rhyme-OldFill | 930 (32) | 832 (32) | 521 (14) | 493 (14) | 750 (21) | 558 (12) | 938 (29) |
| Rhyme-NewFill | 893 (30) | 722 (23) | 497 (13) | 488 (14) | 741 (22) | 555 (13) | 1007 (28) |
| NoRhyme-OldFill | 825 (27) | 701 (21) | 501 (12) | 485 (12) | 768 (24) | 553 (13) | 1019 (32) |
| NoRhyme-NewFill | 839 (26) | 726 (21) | 500 (14) | 479 (13) | 719 (20) | 560 (13) | 1037 (32) |

**Table 6**

Summary of statistical effects by region for Experiment 2. Asterisks indicate significant effects. Absolute values of *t* > 2 were considered significant at a p < .05 level. All models contained random slopes for all fixed effects and their interaction by subject and item except for Region 1, where the items term contained random slopes for the fixed effects, but not their interaction.

| | | Estimate | S.E. | t value |
|---|---|---|---|---|
| Region 1 | **(Intercept)** | 879.35 | 48.4 | 18.166* |
| | **Filler** | 5.574 | 28.8 | 0.193 |
| | **List** | 70.44 | 28.4 | 2.479* |
| | **Filler × List** | 42.75 | 48.7 | 0.878 |
| Region 2 | **(Intercept)** | 751.11 | 38.3 | 19.609* |
| | **Filler** | 35.63 | 22.7 | 1.570 |
| | **List** | 61.29 | 21.9 | 2.798* |
| | **Filler × List** | 115.37 | 45.8 | 2.520* |
| Region 3 | **(Intercept)** | 506.87 | 15.7 | 32.36* |
| | **Filler** | 13.02 | 17.2 | 0.76 |
| | **List** | 6.62 | 13.4 | 0.50 |
| | **Filler × List** | 23.49 | 25.6 | 0.92 |
| Region 4 | **(Intercept)** | 488.41 | 20.7 | 23.544 |
| | **Filler** | 4.53 | 14.0 | 0.322 |
| | **List** | 6.69 | 13.4 | 0.499 |
| | **Filler × List** | 0.98 | 24.5 | 0.040 |
| Region 5 | **(Intercept)** | 747.76 | 31.9 | 23.435 |
| | **Filler** | 25.05 | 23.7 | 1.057 |
| | **List** | -6.83 | 25.0 | -0.273 |
| | **Filler × List** | -57.80 | 44.3 | -1.306 |
| Region 6 | **(Intercept)** | 558.74 | 17.5 | 31.88 |
| | **Filler** | -5.18 | 14.7 | -0.35 |
| | **List** | -2.19 | 12.4 | -0.18 |
| | **Filler × List** | 7.40 | 31.9 | 0.23 |
| Region 7 | **(Intercept)** | 1007.84 | 49.2 | 20.490 |
| | **Filler** | -45.95 | 29.7 | -1.545 |
| | **List** | -63.82 | 25.5 | -2.501* |
| | **Filler × List** | -29.21 | 55.0 | -0.531 |