Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory

Daniel J. Acheson, Maryellen C. MacDonald *

Department of Psychology, University of Wisconsin-Madison, Madison, WI 53706, United States

ARTICLE INFO

ABSTRACT

Many accounts of working memory posit specialized storage mechanisms for the maintenance of serial order. We explore an alternative, that maintenance is achieved through temporary activation in the language production architecture. Four experiments examined the extent to which the phonological similarity effect can be explained as a sublexical speech error. Phonologically similar nonword stimuli were ordered to create tongue twister or control materials used in four tasks: reading aloud, immediate spoken recall, immediate typed recall, and serial recognition. Dependent measures from working memory (recall accuracy) and language production (speech errors) fields were used. Even though lists were identical except for item order, robust effects of tongue twisters were observed. Speech error analyses showed that errors were better described as phoneme rather than item ordering errors. The distribution of speech errors was comparable across all experiments and exhibited syllable-position effects, suggesting an important role for production processes. Implications for working memory and language production are discussed.

© 2009 Published by Elsevier Inc.

Introduction

Verbal working memory, the temporary maintenance and processing of verbal information, has long been viewed as an important component to word learning (Gathercole & Baddeley, 1989, 1990) and language comprehension (Daneman & Carpenter, 1980). The relationship between working memory and language production has received far less attention, despite the fact that recall of verbal material requires language production (although see Bock, 1996; Ellis, 1980; Jacquemot & Scott, 2006; Jones, Macken, & Nicholls, 2004; Page, Madge, Cumming, & Norris, 2007; Treiman & Danis, 1988). The production-working memory relationship was explored in a recent review by Acheson and MacDonald (2009), who argued that the mechanism for maintaining serial order in verbal working memory may emerge from the language production architecture (hereafter the *language production hypothesis*). In the present study, we investigate this hypothesis in studies of working memory performance, using analytic techniques and behavioral manipulations more typical of language production researchers.

Many researchers' views of the architecture underlying working memory have been shaped by the multi-component model (Baddeley, 1986; Baddeley & Hitch, 1974), in which an attentional control mechanism termed the *central executive* oversees the functioning of two slave systems responsible for the temporary maintenance of verbal (the phonological loop) and visual (the visuospatial sketchpad) information. The phonological loop component of the model is in turn composed of a temporary phonological store whose contents decay with time unless refreshed via an articulatory control process. A number of key phenomena have been used to support the phonological loop concept, among them effects of phonological similarity (Conrad, 1964; Conrad & Hull, 1964; Wickelgren, 1965), word length (Baddeley, Thomson, & Buchanan, 1975), irrelevant sound (Colle & Welsh, 1976; Salame & Baddeley, 1982), and concurrent articulation (Baddeley, Lewis, & Vallar, 1984; Levy, 1971; Murray, 1968). On this view, while participants must engage in language production to com-

* Corresponding author. Fax: +1 608 262 4129.
E-mail address: mcmacdonald@wisc.edu (M.C. MacDonald).

plete the recall task, an independent storage mechanism is responsible for memory maintenance.

One challenge to language-independent, short-term storage comes from findings that show the influence of long-term, linguistic knowledge on putatively short-term recall. For example, words are easier to recall than non-words (Hulme, Maughan, & Brown, 1991), and high frequency words are easier to recall than low frequency ones (Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002). Similarly, concrete words are easier to recall than abstract words (Walker & Hulme, 1999). In addition to these lexical or semantic influences, long-term phonological knowledge affects working memory performance: Non-words with higher phonotactic probabilities (i.e., having higher frequency phonemes in higher frequency combinations) are easier to recall than those with lower phonotactic probabilities (Gathercole, Frankish, Pickering, & Peaker, 1999), and nonwords from more dense phonological neighborhoods (those with more phonologically related words in the language) are recalled better than those from sparse neighborhoods (Roodenrys & Hinton, 2002).

These studies suggest a system which, at the very least, uses linguistic representation that has been acquired over a lifetime to constrain performance on recall tasks. The idea that long-term, linguistic knowledge influences and should be incorporated into accounts of verbal working memory is not a new one (see Allport, 1984; Ellis, 1980; Gupta & MacWhinney, 1997; Hartley & Houghton, 1996). Researchers have used one of the classic findings in verbal working memory, the phonological similarity effect, as a means of testing the functional relationships between language and working memory (Nimmo & Roodenrys, 2004). Past research has demonstrated that memory for the *order* of items sharing phonological features is worse than items which do not share these features, although nonordered memory for the items themselves is generally not affected or may be enhanced (Fallon, Groves, & Tehan, 1999; Gupta, Lipinski, & Aktunc, 2005; Nimmo & Roodenrys, 2004). Critical to the present study, the location of the feature overlap influences people's performance. Both Nimmo and Roodenrys (2004) and Gupta et al. (2005) demonstrated that relative to nonoverlapping lists, when phonological overlap occurs within the rhyme unit of a syllable (e.g. the /æt/ sound in the list *cat, bat, hat, etc.*), people's memory for the order of the stimuli is impaired while their memory for the stimuli improves. When stimuli share the same initial consonant and vowel (e.g. *sane, safe, sake*, etc.) or the same initial and final consonants (e.g. *bought, bet, boot*, etc.), people demonstrate impairments in both item and order memory (Nimmo & Roodenrys, 2004). These studies thus support claims for a role of long-term knowledge in immediate recall by demonstrating that the linguistic structure of the material (in this case, the syllabic structure), is central to how manipulations of phonological similarity affect performance.

Using manipulations of phonological similarity to examine linguistic structure can be difficult, however, as the stimuli used can create an ambiguity in interpreting the linguistic level over which errors are occurring. For example, in the case of serial recall of a sequence of phonologically similar letters (C, B, P, D, etc.), ordering errors can

be interpreted as occurring at the level of whole lexical items (e.g., C, *P*, *B*, D), but they also can be interpreted at the sublexical level, as exchanging of the /b/ and /p/ phonemes in the planned utterance will also result in the C *P B* D error (see Page & Norris, 1998b, for a similar discussion). Analysis of naturally occurring speech errors suggests that ordering effects due to phonological similarity are not mis-orderings of whole words but are more likely have a sublexical source, reflecting errors in phonemes across the same syllable position in different items (Dell, 1984; Shattuck-Hufnagel, 1979). Page et al. (2007) investigated sub-lexical errors in recall using "spoonerized" lures. Spoonerisms occur when an exchange of speech sounds between two words results in the production of real words (e.g., "you've *h*issed my *my*stery lecture" instead of "you've *m*issed my *h*istory lecture;" MacKay, 1970). Thus, when people do make speech errors with such stimuli, there is no ambiguity as to the unit over which the errors is occurring as the error results in the production of an unintended word. Using such lures in a serial recall task, Page et al. (2007) demonstrated that people produced many more errors when the two lures were adjacent to each other than when they were not. The authors concluded that errors due to phonological similarity (in this case, an exchange of speech onsets) likely reflect errors within the speech production system.

In this paper, we hypothesize that viewing the maintenance of serial order information in verbal working memory tasks as a slightly idiosyncratic language processing task can provide insight into working memory processes. Although this view is consistent with recent approaches suggesting that verbal working memory performance is closely linked to language production (Jacquemot & Scott, 2006; Saito & Baddeley, 2004), it differs in that these researchers have assumed "buffers" that are specifically dedicated to short-term maintenance. Alternatively, Page et al. (2007) recently suggested that what has been termed the "phonological loop" can be likened to a list wise, lexical-level production plan. In this account, serial recall amounts to a speech "reproduction" task, and differs from typical production as the source of this plan is not internally generated. This view is quite similar to ours in suggesting that verbal working memory maintenance occurs by maintaining activation over the same levels of representation responsible for normal production, although our emphasis in the present research is at a sublexical level.

Language production researchers have posited a number of subprocesses that are executed in the course of language planning and production. Two of them seem highly relevant to maintaining and recalling an ordered sequence of lexical and sublexical phonological representations. These processes are *lexical retrieval* and *phonological encoding*, in which a to-be-uttered word is translated into a sequence of phonemes prior to articulation (Dell, 1986; Garrett, 1975; Shattuck-Hufnagel, 1979). The term *lexical,* as it is used here and throughout, does not necessarily refer to a word with its associated semantics, but rather, to a whole phonological representation. In this sense, a lexical–phonological representation is dissociable from sublexical representations (e.g. phonemes, phonetic features, etc.). Both lexical and sublexical effects have been analyzed

in corpora of naturally occurring speech errors. Such analyses have provided important empirical evidence about the nature of production processes, including phonological encoding (Fromkin, 1971; Nooteboom, 1969; Shattuck-Hufnagel, 1979), and yield several phenomena that bear a strong resemblance to classic effects in verbal working memory (see Acheson & MacDonald, 2009). If the language production hypothesis is correct, then use of speech error analyses in the context of verbal working memory tasks such as serial recall should reveal important insight into people's performance.

To date, only two attempts have been made to incorporate speech error analyses into serial recall performance. Ellis (1980) examined the extent to which errors in serial recall obey those in normal speech production and demonstrated that errors are more likely to occur between speech elements which share more similar phonetic features (also see Wickelgren, 1965); that errors between consonants are more common than errors between vowels; and that speech sounds are more likely to exchange with each other when they occur within the same syllable position. Treiman and Danis (1988) also examined the extent to which errors in serial recall abided by syllable structure in CVC, CCV and VCC syllables. Across three studies, errors occurred primarily between speech sounds within a list, and they tended to maintain the onset-rhyme distinction within the syllable structure. Thus, errors in verbal working memory abide by what production researchers have called the syllable-position constraint (Dell, 1986). Although both the Ellis (1980) and Treiman and Danis (1988) studies were an important first step in demonstrating the utility of speech error analyses serial recall performance, they lacked the depth of detail typically presented in analyses of naturally occurring speech errors, and they did not include a nonmemory baseline condition.

Our studies expand upon previous research by systematically analyzing the types and distribution of speech errors across items and syllable positions. Such an analysis will, for the first time, provide a detailed taxonomy of the types of errors induced by phonological similarity in memory tasks. The following terminology will be used in this investigation. The level over which an error is occurring (e.g. phoneme, syllable, word) is called the *segment*. Within a given syllable (e.g. *mip*), segments are divided into the initial consonant or consonant cluster (the onset), the middle vowel, and the last consonant(s) (the *coda*, also called the *offset*), where the combination of these last two segments is termed the *rhyme*. We will refer to our stimulus nonwords as *items* to be produced or remembered, so that phonemes, onsets, etc. are components of our stimulus items.

We investigate the sublexical nature of phonological similarity by examining the types of errors people make when producing and remembering tongue twister stimuli. Tongue twisters such as *she sells sea shells by the seashore* are known to produce serial ordering errors both in the process of phonological encoding and in articulation itself (Wilshire, 1999). What often defines a tongue twister is the complex alternation of onset and rhyme coupled with phonetic feature similarity in the onset syllable position. In the example above, the onsets phonemes follow an

ABBA pattern (/ʃ/ /s/ s/ /ʃ/) while the rhymes follow an ABAB pattern (/i/ /ɛl/ /i/ /ɛl/). The logic of using such stimuli is threefold. First, there is general agreement in the production area that tongue twisters primarily elicit phoneme and not whole item errors (Wilshire, 1999). Second, comparisons can be made across stimuli matched for overall phonological overlap, as the same stimuli appear in different orders across tongue twister conditions. Most would agree that a list of more difficult-to-produce stimuli should be harder to recall. However, beyond suggesting that tongue twister lists simply take longer to rehearse (and hence suffer more decay), it is not clear how models of working memory that do not take into account sublexical linguistic structure would account for tongue twister effects. The language production hypothesis, on the other hand, makes explicit why tongue twisters are difficult, as such ordering causes difficulty during phonological encoding. Third, although phoneme errors often result in exchanges between stimuli (e.g. she *shells* sea *sells*), they also result in repetition of the same item (e.g. she *shells* sea *shells*). Item repetitions within a list are very rare in serial recall (Henson, 1998), and this outcome is explicitly built in to many computational models in the form of post-output suppression of entire items (e.g. Burgess & Hitch, 1999; Henson, 1998). Given these mechanisms, item repetition simply should not occur in the list lengths used in this study. In the language production hypothesis, however, item repetition is easily accommodated as a phoneme repetition. Our point is not to argue that item-level maintenance does not occur, nor that post-output suppression of previously spoken material is wrong (quite simply, it can't be). Rather, we argue that the most likely source of an item repetition in tongue twisters is in the repetition of individual phonemes rather than a failure in post-output suppression of an entire item, thus suggesting an important role for sublexical, phonological encoding.

Tongue twisters have been employed in verbal working memory tasks before (e.g. Saito & Baddeley, 2004); however, researchers have not conducted the detailed error analysis provided here. This type of analysis is capable of detecting long-term constraints on the production architecture that may be present in verbal working memory tasks. Furthermore, we manipulate the mnemonic demands in the task by having individuals produce (Experiment 1) and remember (Experiments 2–4) lists of nonwords. Nonwords were chosen as stimuli, as their structure can be tightly controlled within and across lists, and they necessarily lack semantic content, thus influences of semantic processes on serial ordering can be avoided. Unlike in previous studies, the use of a simple production task in Experiment 1 provides a baseline of performance for the following experiments in which participants perform immediate spoken (Experiment 2) or typed (Experiment 3) recall, or serial recognition (Experiment 4).

Drawing on previous research on experimentally-induced and naturally occurring speech errors, the language production hypothesis makes three specific predictions about task performance. First, tongue twister orders should be harder to recall than non-tongue twister orders due to errors in phonological encoding processes. Second, analyses should reveal similar distributions of speech errors

across serial recall and production of tongue twister sequences. In the case of phonological encoding, five basic error types have been classified in the speech error literature: substitutions, exchanges, shifts, additions and omissions (Fromkin, 1971; Garrett, 1975; Nooteboom, 1969; Shattuck-Hufnagel, 1979; examples are provided in Appendix A). Given that nature of the stimuli, the primary error we predict is a contextual substitution (Wilshire, 1999), in which a target segment (the element that was intended to be uttered) is replaced by an intruding segment from elsewhere in the utterance. Third, we predict that errors due to phonological similarity reflect phoneme ordering, not errors in ordering whole items, thus phoneme substitutions will result in repetition of the same items due to perseveration (i.e., repetition of material already spoken) or anticipation (i.e. production of upcoming material) of phonemes within the list. Finally, the distribution of phoneme errors across syllables should reflect structural and distributional constraints on the production system. Thus, phoneme substitution errors should abide by syllable-position constraints, and the majority of these errors should occur at the onset consonant(s) (Shattuck-Hufnagel, Keller, & Gopnik, 1987).

Experiment 1: rapid, paced reading of tongue twisters

Before we embark on analysis of speech errors in the context of verbal working memory tasks, it is useful to have a nonmemory baseline measure of production errors for the same items that will be used in the memory studies. We therefore developed a list of items that would be likely to induce speech errors and instructed producers to read each list of items repeatedly in a rapid, paced manner. Such a technique has been shown to produce speech errors reliably in tongue twister lists (Wilshire, 1999).

Method

Participants

Sixteen native English speakers (10 female) participated in this study for credit in an introductory psychology class at the University of Wisconsin, Madison. Their age ranged from 18 to 25 ($M = 19.1$, $SD = 1.70$). All had normal or corrected-to-normal vision.

Materials

The stimuli consisted of 28 lists of four CVC or CVCC nonwords, all of which are listed in Appendix B. Individual lists were composed of two pairs of items sharing a similar vowel and coda (i.e., rhyme), but whose onsets and codas differed by at most two phonetic features (e.g. place, voicing, frication, etc.). For instance, in the list *shif sheev sif seev*, the first and third items (*shif, sif*) share a rhyme (/ɪf/) but differ in their onset phoneme (/ʃ/ and /s/, respectively).

Each set of four nonwords were arranged in a tongue twister and a non-tongue twister pattern. Whereas rhymes consistently followed an ABAB pattern in both conditions (i.e. / ɪf / /iv/ / ɪf / /iv/), tongue twisters were created by varying the pattern of onsets. Non-tongue twisters were composed of an AABB onset pattern (e.g., *shif sheev sif seev*). Tongue twisters were created by switching the second and

fourth onset of the non-tongue twister pattern, creating an ABBA onset pattern (e.g., *shif seev sif sheev*). A common example of this pattern of onsets and rhymes comes in first four words of the well-known English tongue twister *she sells sea shells by the seashore*. The alternating pattern of ABBA onsets with ABAB rhymes has been shown to produce speech errors consistent with tongue twisters (Wilshire, 1999).

Procedure

Stimuli were presented on a computer screen in white font on a black background and printed in 24-point lower-case characters using the Courier New font. Participants first performed a nonword reading task, in which every nonword in the entire stimulus set was presented individually in the middle of screen at a rate of 1 item per second, in random order, and participants read the items aloud. This task allowed us to assess how individuals read the nonwords without the temporal demands and phonological similarity of the list reading task.

After reading the individual nonwords, participants completed the list reading task. When a participant initiated a trial by pressing a computer key, a list of nonwords was displayed on the computer screen, with the four items arranged in a single row. The list remained on screen for two seconds before a response prompt was given; this two second viewing interval allowed participants to read the list and prepare for the paced reading task. After the two second interval, a beep sounded, and the first item on the list was underlined and changed from white to yellow on the screen, cuing the participant to read this item aloud. Subsequent items were signaled with the yellow + underline cue in turn, so that participants were prompted to read a new item every 550 ms. Participants read through each list of four nonwords five times in each trial. Responses were audio-recorded for later analysis. Participants were given a break halfway through the task (14 trials).

Participants read lists of tongue twisters and non-tongue twisters in blocks of seven consecutive trials. Two lists were created across participants such that the same seven items presented as tongue twisters for one group of participants were presented as non-tongue twisters for the other group.

Transcription

Recordings were digitized at a sampling rate of 22 kHz prior to transcription. Two trained individuals transcribed the participant responses using the phoneme conventions from the CMU Pronouncing Dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict). The transcribers were instructed to err on the side of being conservative in their transcription, that is, to code all doubtful cases as correct rather than errors. For example, transcribers were instructed to code as correct any phonemes that were ambiguous between a correct and incorrect response, and they coded participant's self-corrections (e.g. "shi…sif") as correct. The agreement across transcribers was very high. Of the four participants transcribed by both, their transcriptions agreed 92.4% of the time.

Error coding

Error coding was conducted by comparing the spoken responses to a target list generated from the initial reading of the four nonwords in isolation. At times, participants would misread the nonwords initially (e.g. reading /sɪv/ as /sɪv/), but correct themselves on the first trial of the paced list reading task. Target transcriptions were adjusted to correspond to the most consistent way in which participants read each of the nonwords, typically the pronunciation given in their initial reading.

Three main errors were coded: additions, omissions and substitutions. Although we also monitored for shifts, these errors never occurred. Substitutions were further classified as exchanges (i.e. a dual substitution between two list items) and repetitions of items within a list (see Appendix F). Additions, omissions and substitutions were coded both at the level of individual phonemes and at the level of entire items; exchanges and repetitions were coded only at the level of entire items. Error coding was automated with a computer script. Individual items were coded for only one major type of error, thus an item could not be coded as both a substitution and an addition. Based on their proportions in naturally occurring speech errors, we anticipated that there would be more substitutions than omissions, and more omissions than additions; therefore, additions were coded first, followed by omissions, then substitutions. Examination of the error data revealed that only a small percentage of total erroneous responses (<3%) would have been jointly classified. Although this technique of error coding may lose some sensitivity in coding for every single type of error produced by an individual, the overall scheme produced distributions of speech errors commensurate with other researchers (e.g. Meijer, 1997; Wilshire, 1999).

Substitutions were coded for whether they were contextual (i.e., those in which the substituted segment was contained in the current list) or noncontextual (i.e., not in the current list). Contextual substitutions were coded for the item and position from which they came. In that a common type of substitution in tongue twisters is to create an exchange of phonemes (a dual substitution as in *she shells sea sells*), some phoneme contextual substitutions were also coded as item contextual substitutions. At times, substitutions resulted in repetition of the same item within the list (e.g. *shif sheev shif seev*), and these repetitions were coded for whether they were anticipatory (i.e., occurring in an earlier list position) or perseveratory (i.e., occurring later).

Only trials in which participants correctly encoded the target nonwords during the initial reading task were included for analysis. Thus, across pairs of words, participants were required to have been consistent in how they read the onsets, vowels and codas of the matched sets of words (e.g., *shif* and *sif*). This was done to ensure that when an error was made, it was not due to an error in how participants encoded the items. Because targets were transcribed "leniently" in this experiment, by re-transcribing trials in which participants corrected themselves during the first reading of the entire list, no data were excluded in the following analyses.

Statistical analysis

In order to compare tongue twister and non-tongue twister lists, both subjects (F1) and items (F2; Clark, 1973) were treated as random effects. Although the former analysis is typically employed in working memory research, the latter, which is more common in psycholinguistic research, provides a direct test of the tongue twister manipulation in that tongue twister and non-tongue twister lists were composed of the exact same items in different orders. Such an analysis is a useful addition to a subject-based analysis in that the tongue twister manipulation can be examined across item lists whose phonological similarity is matched.

Results

The types of speech error observed were commensurate with our predictions about tongue twisters in speech production. Tongue twister orders produce more errors than non-tongue twister orders; errors were primarily contextual and the distribution of these errors abided by syllable-position constraints. Results presented were consistent across the lists used between participants.

Error analysis

The overall error rate in the present study was very small (<10% of total responses), which is not surprising given that participants were reading each of the nonwords and could see the entire list as they read. A breakdown of the total number of each type of error classified in this experiment is provided in Appendix C. In order to minimize type I error in the univariate analyses below, a 6 (Error Type) × 4 (List Position) × 2 (Tongue Twister) MANOVA was conducted on the mean proportion of errors as a function of the total items produced (both correct and incorrect). As with subsequent MANOVAs, Pillai's trace criterion was used to evaluate multivariate effects. Results are presented in 1. The MANOVA revealed a significant multivariate main effect of Error Type, as well as multivariate interactions of Tongue Twister × List Position, and Tongue Twister × Error Type. These results demonstrate a strong multivariate effect of the tongue twister manipulation, which varied as a function of serial position and error type. The univariate tests for each error type below decompose these interactions, with results listed in Table 2.

Table 1
Results of the multivariate analysis of variance (MANOVA) for error scoring in Experiments 1, 2 and 3.

	Experiment 1	Experiment 2	Experiment 3
Tongue twister (TT)	$F(1,15) = 14.79$	$F(1,29) = 133.68$	$F(1,19) = 71.08$
Position (P)	$F(3,13) = 0.69$	$F(3,27) = 12.29$	$F(3,57) = 15.21$
Error type (E)	$F(5,11) = 5.04$	$F(5,25) = 39.99$	$F(5,15) = 20.45$
TT × P	$F(3,13) = 4.64$	$F(3,27) = 5.59$	$F(3,17) = 2.87$
TT × E	$F(5,11) = 4.26$	$F(5,25) = 12.24$	$F(5,15) = 9.98$
P × E	$F(1,15) = 0.18$	$F(15,15) = 18.63$	$F(5,15) = 2.69$
T × P × E	$F(1,15) = 1.83$	$F(15,15) = 6.22$	$F(5,15) = 2.59$

Note: Bold font indicates significant multivariate effects ($p < .05$ using Pillai's trace criterion); Italics indicates $.05 < p < .10$.

Table 2
Results of univariate ANOVAs for Repeated Reading of Tongue Twisters (Experiment 1)[a].

	Tongue twister	Position	Tongue twister × Position
Speech errors			
Substitutions			
Noncontextual			
Phoneme	$F1(1,15) = 2.85, p > .1$	$\textbf{F1(3,45) = 2.90, p < .05}$	$\textbf{F1(3,45) = 3.577, p < .05}$
	$F2(1,27) = 1.25, p > .2$	$F2(3,81) = 1.14, p > .30$	$F2(3,81) = 2.61, p = .057$
	$\min F'(1,42) = 0.87, p > .6$	$\min F'(3,123) = 0.82, p > .4$	$\min F'(3,124) = 1.51, p > .20$
Contextual			
Phoneme	$\textbf{F1(1,15) = 10.26, p < .01}$	$F1(3,45) = 0.45, p > .7$	$\textbf{F1(3,45) = 6.75, p < .001}$
	$\textbf{F2(1,27) = 11.13, p < .01}$	$F2(3,81) = 0.52, p > .6$	$\textbf{F2(3,81) = 6.56, p < .001}$
	$\textbf{min}\,\textbf{F'(1,38) = 5.34, p < .05}$	$\min F'(3,117) = 0.24, p > .7$	$\textbf{min}\,\textbf{F' (3,117) = 3.33, p < .05}$
Item	$\textbf{F1(1,15) = 12.90, p < .01}$	$F1(3,45) = 0.77, p > .5$	$\textbf{F1(3,45) = 5.17, p < .01}$
	$\textbf{F2(1,27) = 17.16, p < .001}$	$F2(3,81) = 1.04, p > .3$	$\textbf{F2(3,81) = 6.37, p < .01}$
	$\textbf{min}\,\textbf{F'(1,35) = 7.36, p < .02}$	$\min F'(3,104) = 0.44, p > .5$	$\textbf{min}\,\textbf{F'(3,112) = 3.01, p < .05}$
Omissions			
Phoneme	$F1(1,15) = 0.54$	$F1(3,45) = 2.43, p < .08$	$F1(3,45) = 0.74, p > .5$
	$F2(1,27) = 1.09, p > .3$	$F2(3,81) = 1.20, p > .3$	$F2(3,81) = 1.08, p > .3$
	$\min F'(1,30) = 0.36, p > .7$	$\min F'(3,126) = 0.8, p > .6$	$\min F'(3,101) = 0.44, p > .7$
Item	$F1(1,15) = 2.01, p > .1$	$F1(3,45) = 1, p > .4$	$F1(3,45) = 0.4, p > .7$
	$\textbf{F2(1,27) = 8.64, p < .01}$	$F2(3,81) = 0.84, p > .4$	$F2(3,81) = 0.35, p > .8$
	$\min F'(1,22) = 1.63, p > .2$	$\min F'(3,121) = 0.46, p > .5$	$\min F'(3,120) = 0.19, p > .7$
Additions			
Phoneme	$F1(1,15) = 0.05, p > .8$	$\textbf{F1(3,45) = 3.39, p < .03}$	$F1(3,45) = 0.56, p > .6$
	$F2(1,27) = 0.14, p > .7$	$F2(3,81) = 1.58, p > .2$	$F2(3,81) = 0.90, p > .4$
	$\min F'(1,26) = 0.04, p > .8$	$\min F'(3,125) = 1.08, p > .3$	$\min F'(3,97) = 0.35, p > .6$

[a] Bold font indicates a significant difference ($p < .5$); Italic font indicates a trend towards significance ($.05 < p < .10$).

Substitutions. Substitutions in the study were classified as either contextual or noncontextual, and at the level of items and phonemes. Fig. 1 contains a graph of the mean proportion of both of these types of phoneme substitutions relative to the total number of items spoken for each level of the tongue twister manipulation. For this and all subsequent figures, error bars reflect the standard error of the mean. It is clear from Fig. 1 and Appendix C that contextual substitutions outnumbered noncontextual ones.

Contextual substitutions at the level of individual phonemes were more common in the tongue twister than non-tongue twister condition. The results of a 4 (List Position) × 2 (Tongue Twister condition) ANOVA on the mean proportion of phoneme contextual substitutions revealed a main effect of Tongue twister condition, an interaction between List Position and Tongue Twister condition, but no main effect of List Position. The proportion of phoneme contextual substitutions was greater for tongue

twisters relative to non-tongue twisters (mean difference 2.0%, 95% CI ± 1.1%). The interaction emerged from crossover between the tongue twister and non-tongue twister conditions (non-tongue twister > tongue twister) at the second serial position. A very large percentage of the phoneme contextual substitutions could also be classified as item contextual substitutions (~85%). Thus, the overwhelming majority of contextual substitutions in the present experiment were across items with similar rhyme sounds, resulting in a number of repetitions or exchanges of items within the list. Appendix F contains a breakdown of these item contextual substitutions into exchange errors and repetitions. In this instance, the majority of substitution errors at the item level were repetitions of the same items within a list (~70%), and these repetitions tended to be perseveratory. Thus, people had a tendency to repeat items they had already said when they made a repetition error.

Relative to contextual substitutions, phoneme noncontextual substitutions were far fewer (overall proportion <1%), and showed virtually no effect of the tongue twister manipulation. The same ANOVA as above revealed an inconsistent main effect of List Position was observed. There was no main effect of Tongue Twister condition, and an inconsistent interaction between List Position and Tongue Twister condition. Based on these results it would be safe to conclude, however, that the tongue twister effect was observed for contextual substitutions only.

As noted in the introduction, one of the tests of whether language production processes underlie serial ordering in verbal working memory comes from the examination of the syllable-position constraint. Thus, beyond addressing whether the types of substitutions individuals make are occurring at the level of individual phonemes or whole items, we can also investigate the extent to which errors
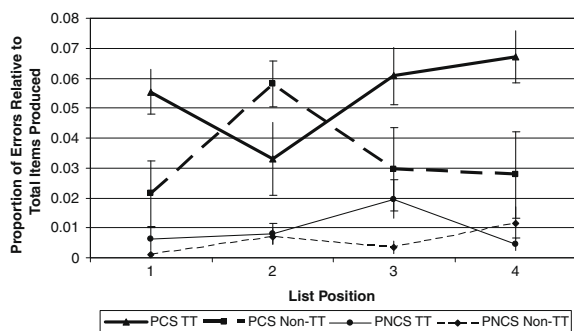

Fig. 1. Relative proportion of items with phoneme contextual substitutions (PCS) and phoneme noncontextual substitutions (PNCS) as a function of List Position for rapid reading of tongue twisters (TTs) in Experiment 1.

adhered to the serial position within an item, and whether the onset syllable position was most susceptible to substitution, as it is in naturally occurring speech errors. Table 3 shows the total number of phoneme contextual substitutions for both the tongue twister and non-tongue twister conditions at each position of the nonword (i.e., onset, vowel and coda). It is clear from the table that the syllable-position constraint was present in the current experiment, in that participants always substituted onsets with onsets, vowels with vowels, and codas with codas. The table also reveals that by far the most prevalent type of phoneme substitution was an onset substitution, as predicted from previous analyses of naturally occurring speech errors (Shattuck-Hufnagel et al., 1987).

Additions and omissions. Additions and omissions were very rare in this experiment (combined < 1% of total responses). Appendix C contains the total number of additions and omissions at both the phoneme and item level. Beyond their overall rarity, there was virtually no effect of the tongue twister manipulation.

Discussion

The results from Experiment 1 established that when no memory is required of participants, the tongue twister manipulation affects only the likelihood of committing a contextual substitution, with no concurrent effects on non-contextual substitutions, additions or omissions. The nature of the tongue twister effect thus explains the multivariate interaction between Tongue Twister and Error Type. Contextual substitutions were also responsible for the Tongue Twister-by-Position interaction. The vast majority of contextual substitutions could be classified as occurring both at the level of individual phonemes and whole items. This is not surprising given that participants viewed the entire list of nonwords as they read them rapidly. The errors also strongly obeyed the serial position constraints previously observed in spontaneous speech errors and studies employing tongue twister stimuli. These results thus establish a baseline of the pattern of errors that can be attributed to purely language production (and reading). With this baseline in hand, Experiment 2 examined production of these same materials in a memory task.

Experiment 2: immediate serial recall of tongue twisters – spoken response

Method

Participants

Thirty-four native English speakers (18 female) participated in this study for credit in an introductory psychology class at the University of Wisconsin, Madison. Their age ranged from 18 to 20 ($M = 18.8$, $SD = 0.56$). All had normal or corrected-to-normal vision. Due to malfunctions in recording equipment, data from four participants were lost.

Materials

Materials were the same as in Experiment 1.

Procedure

Participants read the lists of nonwords on a computer screen, which were presented in white font on a black background and printed in 24-point lowercase characters using the courier new font. Participants initiated each trial with a keypress after which a fixation cross was presented 1500 ms. In the encoding part of the trial, the four nonwords were presented one-at-a-time on the screen, and participants read each aloud as it appeared. Each nonword was presented for 750 ms, with a fixation cross displayed for 250 ms, so that the reading rate was one word every 1000 ms. After the offset of the last nonword's fixation cross, participants saw a yellow question mark, which was their cue to recall the nonwords in the order in which they were presented. Participant responses were recorded digitally at sampling rate of 22 kHz. After recall, participants pressed a key to move on to the next trial. Participants were given a break halfway through the experiment.

Transcription

Transcription was performed by one transcriber using the same criteria as in Experiment 1. As a reliability check, three participants' responses were transcribed by a second transcriber, yielding 94.6% agreement on the transcriptions.

Error coding and serial scoring

Error coding was conducted in the same way as in Experiment 1. In this experiment, target utterances were based on how participants read aloud each of the nonwords during encoding just before the recall of the items. Items were also scored according to whether they were recalled correctly or not. Unless otherwise noted, strict scoring criteria were used such that an item was scored as correct only when it was recalled in the correct serial position; this scoring criterion is the most common one used in studies of serial recall.

Only trials in which participants correctly encoded (read aloud) the target nonwords were included for analysis. This was done to ensure that when an error was made,

Table 3
Positions over which phoneme contextual substitutions (PCS) were made within a syllable during repeated reading of tongue twisters (Experiment 1)[a].

Position from which PCS came	Position of PCS			
	Onset	Vowel	Coda	Total
Tongue twister Condition				
Tongue twister				
Onset	**217**	0	0	217
Vowel	0	**17**	0	17
Coda	0	0	**18**	18
Total	217	17	18	252
Non-tongue twister				
Onset	**125**	0	0	125
Vowel	0	**9**	0	9
Coda	0	0	**27**	27
Total	125	9	27	161

[a] Bold font indicates the highest count for the error in each row.

it was not due to an error in how participants initially encoded the items. A total of 15% of the trials were excluded from the following analyses as a result incorrect encoding.

Results

Analyses typical of both the working memory and production literatures were conducted. Overall recall performance is presented first, followed by an analysis of the proportion and types of errors that participants made. Table 4 contains the results of the univariate analyses discussed below. (see Table 5).

Recall performance

Recall accuracy reflected robust effects of the tongue twister manipulation. Fig. 2 shows mean accuracy for each tongue twister condition as a function of list position. A 4 (List Position) × 2 (Tongue Twister) ANOVA revealed a main effect of List Position, a main effect of the Tongue Twister manipulation and an interaction between List Position × Tongue Twister. Examination of Fig. 2 reveals that tongue twister lists were harder to recall than non-tongue twister lists (mean difference 27%; 95% CI ± 3.9%). The position effect is explained by the superior recall at position 1 relative to position 3, with positions 2 and 4 falling somewhere in between. The interaction between position and the tongue twister manipulation can be understood by the larger difference in the tongue twister manipulation at positions 1 and 3 relative to positions 2 and 4.

Relative to Experiment 1, the overall error rate was significantly higher (see Table 6). The overall recall analysis

Table 5
Positions over which phoneme contextual substitutions (PCSs) were made within a syllable during spoken recall of tongue twisters in memory (Experiment 2).[a]

Position from which PCS came	Position of PCS			
	Onset	Vowel	Coda	Total
Tongue twister condition				
Tongue twister				
Onset	**315**	0	33	348
Vowel	0	**60**	0	60
Coda	0	0	**94**	94
Total	315	60	127	502
Non-tongue twister				
Onset	**158**	0	22	180
Vowel	0	**33**	0	33
Coda	0	0	**51**	51
Total	158	33	73	264

[a] Bold font indicates the highest count for the error in each row.

reveals a large effect of the tongue twister manipulation but provides only a coarse view of the particular mistakes underlying participant performance. The error analysis below fills in this detail through classifying the quantity and types of errors individuals made, while providing tests of the language production hypothesis.

Error analysis

A breakdown of the total number of each type of error classified in this experiment is provided in Appendix D. As in Experiment 1, a 6 (Error Type) × 4 (List Position) × 2 (Tongue Twister) MANOVA was conducted on the mean

Table 4
Results of univariate ANOVAs for spoken recall of tongue twisters (tongue twisters; Experiment 2).[a]

		Tongue twister	Position	Tongue twister × Position
Serial Recall				
	Accuracy	**$F1(1,29) = 77.70, p < .001$**	**$F1(3,87) = 12.11, p < .001$**	**$F1(3,87) = 5.61, p < .01$**
		$F2(1,27) = 92.32, p < .001$	**$F2(3,81) = 11.81, p < .001$**	**$F2(3,81) = 7.74, p < .001$**
		$minF'(1,56) = 42.19, p < .001$	**$minF'(3,168) = 5.98, p < .001$**	**$minF'(3,165) = 3.25, p < .05$**
Speech Errors				
Substitutions				
Noncontextual				
	Phoneme	$F1(1,29) = 3.50, p < .08$	**$F1(3,87) = 4.82, p < .01$**	$F1(3,87) = 0.80, p > .5$
		$F2(1,27) = 4.54, p < .05$	**$F2(3,81) = 5.24, p < .01$**	$F2(3,81) = 0.57, p > .6$
		$minF'(1,56) = 1.98, p > .16$	*$minF'(3,168) = 2.51, p < .06$*	$minF'(3,161) = 0.33, p > .7$
Contextual				
	Phoneme	**$F1(1,29) = 34.45, p < .001$**	**$F1(3,87) = 3.87, p < .02$**	**$F1(3,87) = 6.43, p < .01$**
		$F2(1,27) = 38.96, p < .001$	**$F2(3,81) = 6.43, p < .001$**	**$F2(3,81) = 4.53, p < .01$**
		$minF'(1,56) = 18.28, p < .001$	*$minF'(3,161) = 2.42, p < .06$*	**$minF'(3,161) = 2.66 p < .05$**
	Item	**$F1(1,29) = 31.16, p < .001$**	**$F1(3,87) = 3.31, p < .03$**	**$F1(3,87) = 7.02, p < .001$**
		$F2(1,27) = 32.85, p < .001$	**$F2(3,81) = 5.77, p < .001$**	**$F2(3,81) = 5.35, p < .01$**
		$minF'(1,56) = 15.99, p < .02$	$minF'(3,159) = 2.10, p < 0.1$	**$minF'(3,163) = 3.05, p < .05$**
Omissions				
	Phoneme	$F1(1,29) = 0.22, p > .6$	**$F1(1,87) = 5.93, p < .001$**	$F1(3,87) = 0.17, p > .92$
		$F2(1,27) = 0.01 p > .9$	*$F2(3,81) = 2.37, p < .08$*	$F2(3,81) = 0.14, p > .9$
		$minF'(1,29) = 0.01, p > .9$	*$minF'(3,138) = 2.37, p < .08$*	$minF'(3,168) = 0.09, p > .9$
	Item	**$F(1,29) = 8.32, p < .01$**	**$F1(3,87) = 6.19, p < .001$**	**$F1(3,87) = 2.79, p < .05$**
		$F2(1,27) = 46.81, p < .001$	**$F2(3,81) = 29.78, p < .001$**	**$F2(3,81) = 10.43, p < .001$**
		$minF'(1,39) = 7.06, p < .02$	**$minF'(3,121) = 5.12, p < .01$**	*$minF'(3,130) = 2.20, p < .10$*
Additions				
	Phoneme	**$F1(1,29) = 4.85, p < .04$**	$F1(3,87) = 1.08, p > .3$	$F1(3,87) = 1.59, p > .2$
		$F2(1,27) = 2.37, p > .10$	$F2(3,81) = 0.76, p > .5$	$F2(3,81) = 2.10, p > .1$
		$minF'(1,49) = 1.59, p > .20$	$minF'(3,160) = 0.44, p > .7$	$minF'(3,166) = 0.9, p > .4$

[a] Bold font indicates a significant difference ($p < .05$); Italic font indicates a trend towards significance ($.05 < p < .10$).
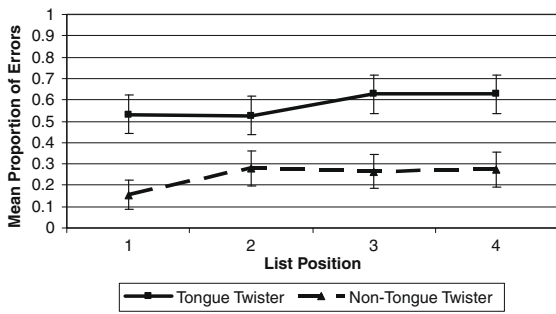
Fig. 2. Mean proportion of errors as a function of list position for the tongue twister manipulation in spoken recall (Experiment 2) using strict serial scoring criteria.

Fig. 3. Relative proportion of items with phoneme contextual substitutions (PCS) and phoneme noncontextual substitutions (PNCS) as a function of list position for spoken recall of tongue twisters (TTs) in Experiment 2.

proportion of errors as a function of the total items produced. Results are presented in Table 1. Significant multivariate main effects were observed for Tongue Twister condition, List Position and Error Type, and significant multivariate interactions for all combinations of these factors. The multivariate test thus reveals multivariate effects of the tongue twister manipulation on error rates, modulated both by list position and error type.

Substitutions. Fig. 3 contains a graph of the mean proportion of both types of phoneme substitutions for each level of the tongue twister manipulation. The first test of the language production hypothesis is to test the prediction that contextual substitutions outnumber noncontextual ones. Based on this graph and the descriptive statistics in Appendix D, it is clear that this prediction is supported. In comparing across the tongue twister conditions for items containing a phoneme contextual substitution, a 4 (List Position) × 2 (Tongue Twister) ANOVA revealed a significant main effect of List Position, the Tongue Twister
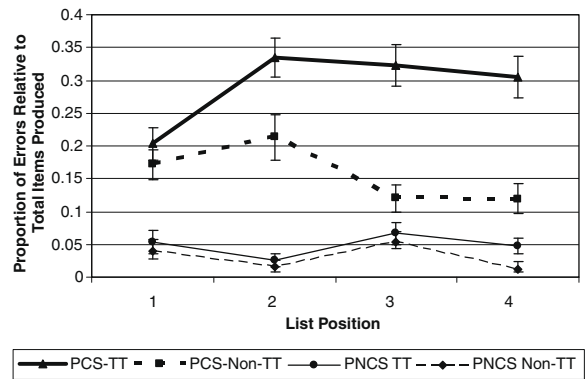
manipulation and a List Position × Tongue Twister interaction. Examination of Fig. 3 reveals that the position effect stems from fewer errors at positions 2 relative to the other positions. The number of items containing a phoneme contextual substitution was substantially higher for the tongue twister relative to the non-tongue twister condition (mean difference 13%; 95% CI ± 4%). Finally, the difference between the tongue twister and non-tongue twister conditions was larger at the third and fourth positions relative to the first and second.

As with the no memory baseline, the vast majority (88%) of phoneme contextual substitutions resulted in production of an item from the list. Appendix F reports these item contextual substitutions as exchanges and repetitions. While the majority (68%) of these item-level substitution errors could be classified as an exchange error, a

Table 6
95% confidence intervals comparing error rates across different experimental conditions and error types.[a]

		Spoken recall – reading	Typed recall – reading	Typed recall – spoken recall	Recognition – reading	Spoken recall – recognition	Typed recall – recognition
Serial recall Error rate	Difference	**0.50** **95% CI ± 0.07**	**0.57** **95% CI ± 0.08**	0.07 95% CI ± 0.10	0.05 95% CI ± 0.06	**0.45** **95% CI ± 0.08**	**0.52** **95% CI ± 0.09**
Substitutions							
Noncontextual							
Phoneme	Difference	**−0.04** **95% CI ± 0.03**	**0.10** **95% CI ± 0.05**	**0.07** **95% CI ± 0.05**			
Item	Difference	0.00	0.00	0.00			
Contextual							
Phoneme	Difference	**0.18** **95% CI ± 0.06**	**0.17** **95% CI ± 0.07**	−0.01 95% CI ± 0.08			
Item	Difference	**0.16** **95% CI ± 0.06**	**0.10** **95% CI ± 0.06**	−0.06 95% CI ± 0.07			
Omissions							
Phoneme	Difference	0.01 95% CI ± 0.02	0.01 95% CI ± 0.02	0.00 95% CI ± 0.02			
Item	Difference	**0.07** **95% CI ± 0.03**	**0.04** **95% CI ± 0.03**	−0.03 95% CI ± 0.05			
Additions							
Phoneme	Difference	**0.04** **95% CI ± 0.03**	**0.05** **95% CI ± 0.04**	0.01 95% CI ± 0.05			
Item	Difference	0.00	0.00	0.00			

[a] Bold font indicates a significant difference between error rates ($p < .05$).

very large percentage (32%) resulted in repetition of the same item, thus confirming one of the predictions of the language production hypothesis. Further examination of these repetition errors revealed a slight perseveratory bias for the tongue twister condition, and an equal distribution of anticipations and perserverations in non-tongue twisters.

Noncontextual substitutions, though fewer in number than contextual ones, also showed a small effect of the tongue twister manipulation. The same ANOVA as above showed a significant main effect of List Position, and Tongue Twister manipulation but no interaction between the two factors. The list position effect can be understood by the relative increase in items containing a phoneme noncontextual substitution at position 4 relative to the other positions. While significant across most analyses (except minF'), the effect of the tongue twister manipulation was small (mean difference 2%; 95% CI ± 0.7%). None of the phoneme noncontextual substitutions in the present experiment could be classified as an item noncontextual substitution.

As a second test of the language production hypothesis, we next examined the extent to which contextual errors adhered to the syllable-position constraint. Table 4 includes a breakdown of the syllable positions over which phoneme contextual substitutions were committed in the present experiment. Assuming an equal probability of phonemes moving within a syllable (i.e. between 25% and 33% depending on whether the syllable is CVCC or CVC), it is clear that the substitutions made were not random. Across both tongue twister conditions, substitutions occurring at the onset (i.e. syllable position 1) came from another onset in the list 88% of the time, whereas the remaining 12% came from the coda. Vowel and coda substitutions came from another vowel or coda, respectively, 100% of the time. As further support of the language production hypothesis, phoneme contextual substitutions showed a very strong onset effect, with 84% of errors occurring at this syllable position; only 7% and 8% of errors occurred at the vowel and coda, respectively.

Additions and omissions. There were very few additions in the present study at either the phoneme or item levels, and the tongue twister manipulation showed virtually no effect. For phoneme additions, the analysis revealed only an inconsistent main effect of the tongue twister manipulation which emerged from a greater proportion of items containing phoneme additions in the tongue twister relative to the non-tongue twister condition (mean difference 0.8%; 95% CI ± 0.3%). Given that lists were always composed of four items, participants never committed an item addition error in this study.

The tongue twister manipulation had little impact on the likelihood that an item contained a phoneme omission. While a significant main effect of List Position was observed, there was no main effect of the Tongue Twister manipulation and an inconsistent interaction.

Unlike omissions at the level of individual phonemes, there were effects of the tongue twister manipulation for item omissions. Analyses revealed a main effect of List Position, a main effect of the Tongue Twister manipulation, and an interaction. Examination of Appendix D reveals that

the position effect emerged from more item omissions at positions 3 and 4 relative to positions one and two. Tongue twister lists were more likely to contain an item omission relative non-tongue twister list (mean difference 6%, 95% CI ± 4%). Finally, the interaction can be understood as a greater effect of the tongue twister manipulation at the third and fourth serial positions relative to the first and second.

Discussion

The results of Experiment 2 show a robust effect of the tongue twister manipulation. Critically, in addition to being present when subjects were treated as a random factor, the same pattern was revealed in a direct comparison across the item with the same amount of phonological overlap (i.e. the F2 analyses). As in the previous experiment in which no memory for items was required, the primary source of the tongue twister effect in serial recall emerged from substitutions. Table 6 contains the direct comparisons of the error rates in all the experiments. Examination of this table reveals that in addition to the overall error rate being higher in spoken recall relative to paced reading, tongue twister effect was observed for noncontextual substitutions, as well as item omissions explaining the multivariate interaction between error type and the tongue twister manipulation. The fact that item omissions showed a tongue twister effect in the present study merits some discussion. We see two complementary explanations as very likely. The first is that errors reflect a failure to fully activate the lexical-level representation of the item. In this sense, omissions reflect memory errors akin to those discussed in models of serial recall, and the fact that these errors tend to occur for later list positions seems to support this conclusion (see Page & Norris, 1998a). The second possible explanation is that omission errors represent the workings of an error monitoring system. Individuals presumably know that there were no repetitions in the stimulus items when they encoded them, yet their speech errors include repetitions. Omission errors in this case might simply represent an instance in which individuals caught themselves from repeating the same items.

Beyond omissions, the overall higher error rate induced by the memory demands relative to previous study had an effect on the extent to which speech errors abided by the typical patterns observed in naturally occurring speech error corpora. For instance, tongue twister effects were observed for noncontextual substitutions, and offset phonemes sometimes substituted for syllable onsets. Research in aphasic patients as well as computational modeling efforts have shown that as the production system breaks down more (i.e., as more errors are produced), the patterns of speech errors begin to depart from normal patterns (see Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). More specifically, a computational model of word production has shown that information decays more quickly, phonological error patterns become more random (Dell, Burger, & Svec, 1997). In the present study, decay of the nonword stimuli being held in memory likely led to a higher error rate and to a few instance in which the syllable-position constraint and bias towards contextual substitution

bias were violated. We return to these points in the general discussion.

Across both tongue twister conditions, over 95% of all recall errors could also be characterized as a speech error of some type—a substitution, omission, or addition. The detailed analysis of these errors confirmed all of the predictions made by the language production hypothesis. Tongue twister orders were harder than non-tongue twister orders. The errors made were primarily contextual substitutions. Substitution errors adhered to the syllable-position constraint and occurred primarily at the onset syllable position. Finally, ordering errors were more reflective of phoneme rather than item substitutions, which resulted in many repetitions of the same items within a list.

Experiment 2's support of the language production hypothesis is limited by an alternative interpretation of the data. Tongue twister effects can emerge not just from phonological encoding, but articulation as well (Wilshire, 1999). Thus, the results of Experiment 2 could plausibly be interpreted as evidence for articulatory errors. Although the impact of articulation during encoding was reduced by removing trials in which items were incorrectly spoken, it might be argued that the effects being observed simply reflect output errors instead of a more central mechanism that underlies serial ordering performance across production and working memory. Experiment 3 was designed to investigate this articulatory alternative.

Experiment 3: immediate serial recall of tongue twisters – typed response

In this experiment, articulatory demands were changed and reduced by having participants type their responses instead of speaking them. Typing, like speaking, is a motor response, but similarities among phonemes are not maintained in the typing modality. For example, the phonemes /p/ and /b/ are highly similar acoustically and are spoken with similar speech gestures, but the letters *p* and *b* require very different motor plans to type them on a keyboard. If the effects in Experiment 2 were purely articulatory, then effects of the tongue twister manipulation should be severely attenuated or abolished across both working memory and error-based analyses. If, however, the serial ordering mechanism underlying performance precedes output, as is the case with phonological encoding in production, then Experiment 3 should yield similar results to those obtained in Experiment 2.

Method

Participants

Twenty native English speakers (eight female) who had not participated in the Experiments 1 or 2 participated in this study for credit in an introductory psychology class at the University of Wisconsin, Madison. Their age ranged from 18 to 19 ($M = 18.9$, $SD = 0.5$). All had normal or corrected-to-normal vision.

Materials

Materials were the same as in Experiment 1.

Procedure

The procedure for the encoding portion of each trial was identical to that in Experiment 2. The recall portion was also identical to the one in Experiment 2, except that participants typed their responses instead of speaking them out loud. Each nonword was typed one-at-a-time during recall; participants were able to see and edit their typed responses. Participants' reading aloud of the items during the encoding phase was recorded digitally at sampling rate of 22 kHz. In addition to analyzing accuracy and error type in the typed responses of participants, total typing time was also examined.

Transcription

Transcription was conducted by translating the typed response into a phonetic code using the same phonetic alphabet as in the first two experiments. Transcription was completed by one individual, and a second transcriber made transcriptions for three participants. Agreement across transcribers was high, with an overall agreement of 95.2%.

Error coding and serial scoring

Error coding was conducted in the same way as in Experiments 1 and 2. As in Experiment 2, target typed responses were based on how participants read each of the nonwords aloud during encoding. Unless otherwise noted, serial recall was scored using the same, strict serial criterion of Experiment 2.

Only trials in which participants correctly encoded the target nonwords were included for analysis, resulting in the exclusion of 11% of the trials.

Results

Results for each of the univariate analyses discussed below are included in Table 7.

Recall performance

Recall accuracy reflected robust effects of the tongue twister manipulation. Fig. 4 contains a graph of the mean accuracy for each tongue twister condition as a function of list position. A 4 (List position) × 2 (Tongue Twister) ANOVA revealed a main effect of List Position, a main effect of the Tongue Twister manipulation and an interaction. Examination of Fig. 4 reveals that the position effect is explained by the superior recall at position 1 relative to position 3, with positions 2 and 4 in between. Recall for tongue twister lists was worse than non-tongue twister ones (mean difference 22%; 95% CI ± 4.6%). Finally, the interaction between List Position and the Tongue Twister manipulation can be understood by the small difference at position 2 to the other positions.

Another means of assessing production-related difficulty comes in the analysis of the total typing time for each item. Fig. 5 contains a graph of the mean response duration for each tongue twister condition as a function of list position. The same ANOVA as above revealed a main effect of List Position, a main effect of the Tongue Twister manipulation but no interaction between the two. The position effect is explained by the linear decrease from the first list

Table 7
Results of univariate ANOVAs for typed recall of tongue twisters (Experiment 3).[a]

		Tongue twister	Position	Tongue Twister × Position
Serial recall				
	Accuracy	$F1(1,19) = 72.26, p < .001$	$F1(3,57) = 13.17, p < .001$	$F1(3,57) = 3.42, p < .03$
		$F2(1,27) = 55.01, p < .01$	$F2(3,81) = 13.45, p < .001$	$F2(3,81) = 5.74, p < .01$
		$minF'(3,133) = 31.23, p < .001$	$minF'(3,133) = 6.65, p < .001$	$minF'(3,116) = 2.14, p < .10$
	Reaction time	$F1(1,10) = 33.80\ p < .001$	$F1(3,57) = 4.20, p < .05$	$F1(3,57) = 1.69, p > .5$
		$F2(1,27) = 7.72, p < .01$	$F2(3,81) = 14.15, p < .001$	$F2(3,81) = 0.90\ p > .1$
		$minF'(1,38) = 6.28, p < .02$	$minF'(3,90) = 3.24, p < .03$	$minF'(3,136) = 0.59, p > .6$
Speech errors				
Substitutions				
Noncontextual				
	Phoneme	$F1(1,19) = 5.82, p < .03$	$F1(3,57) = 0.28, p > .8$	$F1(3,57) = 0.63, p > .6$
		$F2(1,27) = 19.11, p < .001$	$F2(3,81) = 0.3, p > .8$	$F2(3,81) = 0.7, p > .4$
		$minF'(1,30) = 4.46, p < .05$	$minF'(3,132) = 0.14, p > .8$	$minF'(3,131) = 0.33, p > .7$
Contextual				
	Phoneme	$F1(1,19) = 26.82, p < .001$	$F1(3,57) = 12.37, p < .001$	$F1(3,57) = 3.14\ p < .05$
		$F2(1,27) = 34.87, p < .001$	$F2(3,81) = 12.81, p < .001$	$F2(3,81) = 4.16, p < .01$
		$minF'(1,43) = 16.07, p < .001$	$minF'(3,133) = 6.29, p < .001$	$minF'(3,125) = 1.79, p > .15$
	Item	$F1(1,19) = 49.60, p < .001$	$F1(3,57) = 10.25, p < .001$	$F1(3,57) = 4.43, p < .01$
		$F2(1,27) = 59.77, p < .001$	$F2(3,81) = 10.57, p < .001$	$F2(3,81) = 7.12, p < .001$
		$minF'(1,43) = 27.11, p < .001$	$minF'(3,133) = 5.20, p < .01$	$minF'(3,116) = 2.73, p < .05$
Omissions				
	Phoneme	$F1(1,19) = 1.34, p > .2$	$F1(3,57) = 0.83, p > .4$	$F1(3,57) = 1.81, p > .1$
		$F2(1,27) = 0.89, p > .3$	$F2(3,81) = 1.40, p > .2$	$F2(3,81) = 1.98, p > .1$
		$minF'(1,46) = 0.53, p > .4$	$minF'(3,116) = 0.52, p > .6$	$minF'(3,132) = 0.95, p > .5$
	Item	$F1(1,19) = 7.69, p < 0.02$	$F1(3,57) = 2.40, p > .05;$	$F1(3,57) = 0.33, p > .8$
		$F2(1,27) = 38.40, p < 0.001$	$F2(3,81) = 2.96, p < .05$	$F2(3,81) = 0.03, p > .9$
		$minF'(1,27) = 6.41, p < 0.02$	$minF'(3,128) = 1.33, p > .20$	$minF'(3,95) = 0.03, p > .8$
Additions				
	Phoneme	$F1(1,19) = 0.19, p > .6$	$F1(3,57) = 3.80, p < .04$	$F1(3,57) = 0.52, p > .6$
		$F2(1,27) = 0.65, p > .6$	$F2(3,81) = 4.30, p < .01$	$F2(3,81) = 0.47, p > .7$
		$minF'(1,46) = 0.36, p > .6$	$minF'(3,131) = 2.02, p > .10$	$minF'(3,137) = 0.3, p > .6$

[a] Bold font indicates a significant difference ($p < .05$); *Italic* font indicates a trend towards significance ($0.5 < p < .10$).
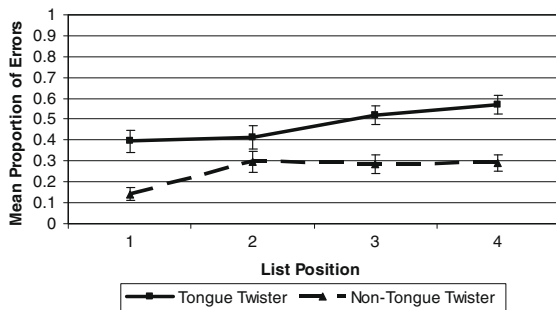

Fig. 4. Mean proportion of errors as a function of list position for the tongue twister manipulation in typed recall (Experiment 3) using strict serial scoring criteria.

Fig. 5. Mean response duration as a function of list position for the tongue twister manipulation in typed recall (Experiment 3) using strict serial scoring criteria.

position to the last position. The main effect of Tongue Twister manipulation was evident in longer response times for tongue twister trials relative to non-tongue twister trials (mean difference 367 ms; 95% CI ± 174 ms). As with Experiment 2, the tongue twister manipulation revealed robust effects. There was no evidence of speed-accuracy tradeoff in the present experiment, in that accuracy was worse and response duration longer for tongue twister lists relative to non-tongue twister lists. The error analysis which follows demonstrates many of the same patterns observed in Experiment 2 with spoken recall.
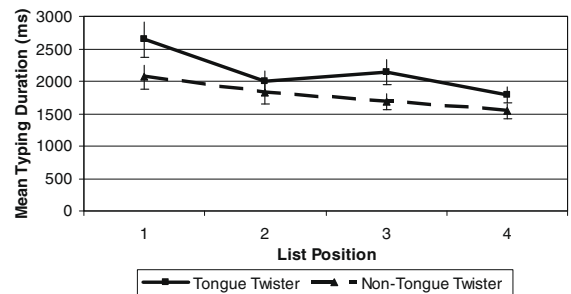
Error analysis

A breakdown of the total number of each type of error classified in this experiment is provided in Appendix E. The three-way MANOVA used in the previous two experiments was repeated here prior to the univariate analyses below. Significant multivariate main effects were observed for Tongue Twister condition, List Position and Error Type. The only significant multivariate interaction was Tongue Twister × Error Type, although the Tongue Twister × List Position effect nearly reached statistical significance. As in the previous experiments, a robust multivariate effect

of the tongue twister manipulation was observed, although this varied as a function of the type of error committed. The univariate analyses below clarify this interaction.

Substitutions. A production-based locus to verbal working memory predicts that contextual substitutions should outnumber noncontextual ones. This prediction was confirmed in this experiment, even though participants typed rather than spoke their responses. Fig. 6 contains a graph of the mean proportion of contextual and noncontextual phoneme substitutions for each level of the tongue twister manipulation and each serial position. In comparing across the tongue twister conditions for items containing a phoneme contextual substitution, a 4 (List Position) × 2 (Tongue Twister) ANOVA revealed a significant main effect of List Position, the Tongue Twister manipulation and a Tongue Twister × List Position interaction. Examination of Fig. 6 and Appendix E reveals that the position effect is understood by the linear increase in items containing phoneme contextual substitutions from positions 1 to 4. The number of items containing a phoneme contextual substitution was again much higher for the tongue twister relative to the non-tongue twister condition (mean difference 14%; 95% CI ± 3.1%). Finally, the difference in the proportion of phoneme contextual substitutions between the tongue twister and non-tongue twister conditions was very small for the second position relative to the other positions.

As with the previous two experiments, the majority of phoneme contextual substitutions (66%) resulted in production of one of the items from the list. Examination of Appendix F reveals that item contextual substitutions in this study were approximately evenly distributed between exchanges (56%) and repetitions (44%), and repetition errors tended to exhibit a perseveratory bias. This large percentage of repetition errors in the present study thus confirms one of the predictions of the language production hypothesis.

Relative to contextual substitutions, noncontextual ones were again fewer in number. An ANOVA revealed a main effect of the Tongue Twister manipulation, but no effect of List Position and no interaction. While significant,
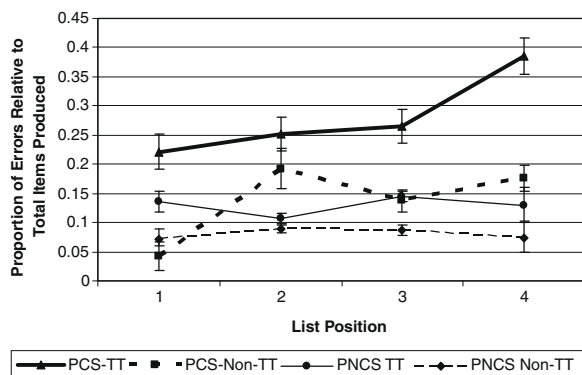
the effect of the tongue twister manipulation was small. Examination of Fig. 6 and Appendix E reveals that tongue twister items containing phoneme noncontextual substitutions only slightly outnumbered the non-tongue twister items (mean difference 4.5%; 95% CI ± 2.4%). None of the phoneme noncontextual substitutions in the present experiment could be classified as an item noncontextual substitution.

As a final test of the language production hypothesis in substitutions, we explored the extent to which phoneme substitutions abided by the syllable-position constraint. Table 8 contains the source of the total number of phoneme contextual substitutions made at each syllable position. Examination of the table reveals that the distribution of the syllable over which substitutions were made was far from random. Across tongue twister conditions, onset substitutions came from other onset 84% of the time, with the remaining 16% coming from the coda. Vowels were substituted with other vowels 99% of the time; the other 1% came from the coda position and likely reflected an error in typing by participants. Coda substitutions came from other codas 100% of the time. Further supporting the language production hypothesis, syllable-onset substitutions were the most common for phoneme contextual substitutions occurring 65% of the time, with vowel and coda substitutions accounted for the remaining 19% and 16% of phoneme contextual substitutions, respectively.

Additions and omissions. There were virtually no additions in the present study at the level of phonemes and none for items, and the tongue twister manipulation showed almost no effect. For phoneme additions, the analysis revealed only a main effect of List Position.

The tongue twister manipulation had no impact on the likelihood that an item contained a phoneme omission. Item omissions, however, showed some effect of the tongue twister manipulation. Although there was an inconsistent main effect of List Position, the Tongue Twister condition was reliable across subjects and items. The interaction between List Position and Tongue Twister condition did not reach significance. Across Tongue Twister conditions, there were significantly more item omissions for

Fig. 6. Relative proportion of items with phoneme contextual substitutions (PCS) and phoneme noncontextual substitutions (PNCS) as a function of list position for typed recall of tongue twisters (TTs) in Experiment 3.

Table 8
Positions over which phoneme contextual substitutions (PCSs) were made within a syllable during typed recall of tongue twisters in memory (Experiment 3).[a]

Position from which PCS came	Position of PCS			
	Onset	Vowel	Coda	Total
Tongue twister condition				
Tongue twister				
Onset	**209**	0	26	235
Vowel	0	**36**	0	36
Coda	0	0	**35**	35
Total	209	36	61	306
Non-tongue twister				
Onset	**69**	0	18	87
Vowel	0	**40**	1	41
Coda	0	0	**34**	34
Total	69	40	53	162

[a] Bold font indicates the highest count for the error in each row.

the tongue twister condition relative to the non-tongue twister condition (mean difference 5%; 95% CI ± 4.0%).

Discussion

The results from Experiment 3 confirm the predictions stemming from the language production hypothesis. As with Experiment 2, the results from Experiment 3 show an effect of the tongue twister manipulation when analyzed from typical approaches taken in the working memory and language production traditions. With the exception of phoneme noncontextual substitutions, the overall error rate is remarkable similar to Experiment 2 (see Table 6). The present experiment shows the same effects as during spoken recall: the tongue twister effects were large for contextual and item omissions, and smaller for noncontextual substitutions, thus explaining the multivariate interaction between tongue twister effect and error type. The results from the present experiment not only demonstrate that the errors in serial ordering due to the tongue twister manipulation are occurring prior to articulation, but that the same production-based constraints are affecting the nature and distribution of the errors made, even when the output modes (speaking vs. typing) are very different.

Although Experiment 3 reduced the articulatory demands on individuals, it is possible that some of the tongue twister effect might still be explained by errors during output itself (e.g. typos) rather than at some earlier stage of output planning (i.e., phonological encoding). In the final experiment, serial output demand is severely reduced through use of a recognition paradigm.

Experiment 4: immediate serial recognition of tongue twisters

Method

Participants

Twenty native English speakers (13 female) participated in this study for credit in an introductory psychology class at the University of Wisconsin, Madison. Their age ranged from 18 to 28 (M = 20.6, SD = 2.5). All had normal or corrected-to-normal vision.

Materials

Materials were generated in the same way as in Experiment 1. Given the need to probe a sufficient number of yes/no responses for each serial position in this recognition task, more items were generated. In addition to the 28 lists used in the previous experiments, an additional 84 lists were developed, resulting in four times as many trials (112) as in the first three experiments. Due to the need to increase the number of stimuli, the phonetic similarity in the onset and coda consonants was not as stringent as in the previous experiments.

Procedure

The encoding portion of each trial was the same as in Experiments 2 and 3, with participants reading each nonword aloud during encoding. After reading the last item, participants were presented with a yes/no recognition probe. Recognition probes assessed not only whether the item had appeared in the list, but whether the item appeared in the correct serial position. This was accomplished by presenting the participant a probe with blank lines indicating items not probed. For instance, in the list *shif sheev sif seev*, participants might see a recognition probe such as: ____ ____ *sif* ____, which queried whether the nonword *sif* had appeared as the third element in the list. In this case, the correct response would have been "yes". "No" responses were generated by transposing items with a similar rhyme sound (e.g. ____ ____ *shif* ____ in the present example). Participants responded by pressing keys marked *yes* or *no* on the keyboard.

As with previous experiments, two lists were generated in which tongue twister orders were presented in their non-tongue twister version across lists. Yes/No responses were counterbalanced within and across lists. Trials were presented randomly.

Scoring

No speech errors were scored in the present experiment. Scoring was either correct or incorrect. As with the previous experiment, reaction time was also recorded. Only trials in which participants correctly encoded (read aloud) the items were included, resulting in the exclusion of 7% of the trials.

Results

Results for each of the univariate analyses discussed below are included in Table 9.

Recognition performance

Serial position curves of mean accuracy for each serial position and each tongue twister condition are presented

Table 9
Results of univariate ANOVAs for serial recognition of tongue twisters (Experiment 4).[a]

	Tongue twister	Position	Tongue twister × Position
Serial recognition			
Accuracy	$F1(1,19) = 39.30, p < .001$	$F1(3,57) = 10.49, p < .001$	$F1(3,57) = 4.78, p < .01$
	$F2(1,111) = 23.86, p < .001$	$F2(3,333) = 8.27, p < .001$	$F2(3,333) = 4.86, p < .01$
	$minF'(1,91) = 14.85, p < .001$	$minF'(3,230) = 4.62, p < .01$	$minF'(3,192) = 2.41, p = .07$
Reaction time	$F1(1,19) = 87.66, p < .001$	$F1(3,57) = 61.32, p < .001$	$F1(3,57) = 48.55, p < .001$
	$F2(1,111) = 1.64, p > .20$	$F2(3,333) = 34.23, p < .001$	$F2(3,333) = 0.43, p > .70$
	$minF'(1,115) = 1.61, p > .2$	$minF'(3,198) = 21.97, p < .001$	$minF'(3,150) = 0.43, p > .5$

[a] Bold font indicates a significant difference ($p < .05$); *Italic* font indicates a trend towards significance ($.05 < p < .10$).

in Fig. 7. Tongue twister effects were smaller than in previous experiments but still present despite the complete absence of serial output. A 4 (List Position) × 2 (Tongue Twister) ANOVA revealed a main effect of List Position, a main effect of Tongue Twister condition, and an interaction between the two. Examination of Fig. 7 reveals that both the position effect and interaction is explained by the near perfect recall at the fourth position across both tongue twister conditions. This result reflects the immediate nature of the recognition probe for the fourth position in this experiment. The main effect of tongue twister condition came from the poorer serial recognition on tongue twister trials relative to non-tongue twister trials (mean difference 9%; 95% CI ± 2.4%).

As with the previous experiment, the tongue twister effect was also present in reaction time data. Fig. 8 presents the mean reaction time (i.e. yes/no response latency) across each tongue twister condition as a function of list position. Missing data in the present analysis were replaced with the mean reaction time collapsed across all serial positions and tongue twister conditions. Tongue twister effects were present again, although they were generally small and less consistent than in the accuracy analysis. An ANOVA revealed a main effect of List Position, an inconsistent main effect of Tongue Twister condition

and an inconsistent interaction, explained by the nearly equal RT across tongue twister condition at the last list position. Examination of Fig. 8 reveals that the difference in RT across tongue twister conditions was fairly modest (mean difference 69 ms; 95% CI ± 97 ms).

Discussion

Results from Experiment 4 show a small effect of the tongue twister manipulation relative to spoken and typed recall (see Table 6). Tongue twisters were fairly consistent in decreasing overall accuracy relative to non-tongue twisters (except for the fourth position). No speed-accuracy tradeoff was observed in the present experiment, in that participants were faster on non-tongue twister trials than tongue twister trials. The RT effect, however, was less consistent in that the direct comparison of items matched for overall phonological overlap (i.e., the F2 analyses) revealed neither a main effect of the tongue twister manipulation nor an interaction with list position. Regardless, these results demonstrate that the tongue twister effect is still present in accuracy measures even when output demands are limited to recognition, thus suggesting a more central mechanism than motor output to the errors in serial ordering performance. This point is further elaborated in the General discussion.

General discussion

The four experiments presented here examined the extent to which serial ordering in verbal working memory relies on the language production architecture. Serial ordering demands were manipulated through use of nonword lists that differed only in the order of items across tongue twister and non-tongue twister conditions. Both mnemonic and output demands were manipulated across experiments to examine the extent to which these factors affected performance. Comparisons across production and working memory were also made through use of analysis techniques utilized by researchers in both traditions in the form of serial position (working memory researchers), speech error and F2/ items (production researchers) analyses. Robust effects of the tongue twister manipulation were observed across all techniques.

We hypothesized that one of the classic findings in verbal working memory, the phonological similarity effect, could primarily be explained by errors within the language production architecture, namely the processes of lexical retrieval and phonological encoding. Based on previous studies of speech errors occurring at these levels of the production architecture, four predictions were generated and tested. First, tongue twister orders should be harder than non-tongue twister orders. Second, error analyses should reveal similar distributions of speech errors across production conditions, and given the stimuli, the primary error made should be a contextual substitution. Third, effects of phonological similarity in tongue twisters primarily stem from ordering errors at the level of phonemes, not items. Thus tongue twister effects should result in repetition of the same item within a list. Finally, errors should reflect long-term learning associ-
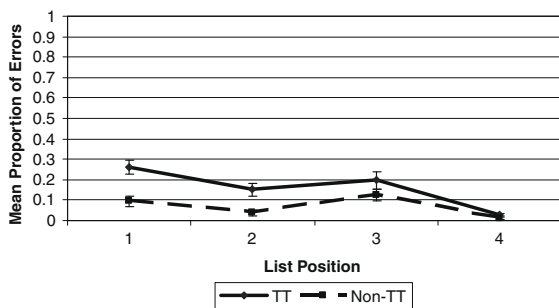

Fig. 7. Mean proportion of errors as a function of list position for the tongue twister (TT) manipulation in serial recognition (Experiment 4) using strict serial scoring criteria.
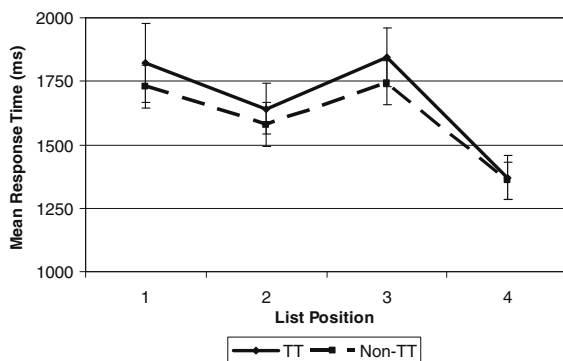

Fig. 8. Mean reaction time as a function of list position for the tongue twister (TT) manipulation in serial recognition (Experiment 4) using strict serial scoring criteria.

ated with syllable structure in production; thus substitution errors should primarily occur across the same syllable position (i.e., the syllable-position constraint), and should primarily occur for the onset of the syllable over the vowel and coda. All of these predictions were confirmed across experiments.

The effect of output demands on the observed error patterns provides important insight into the level at which serial ordering errors are occurring. Whether items were spoken, typed or recognized, tongue twister effects were observed across all experimental conditions, and the distribution of errors was markedly similar. The magnitude of the effect varied somewhat as a function of the task (see Table 6), and we are not claiming that there are no effects of response modality. However, the consistency of the tongue twister result across tasks suggests that most serial ordering errors are occurring at some stage prior to actual output. Thus, similar to the conclusions of other researchers, the present results point to a critical contribution of both phonological encoding and articulation in producing the serial ordering errors observed under conditions of phonological similarity (Gupta & MacWhinney, 1995; Wilshire, 1999).

Before we discuss implications of this research, it is important to note two caveats to the results and interpretation offered here. We have argued that adherence to the syllable-position constraint and the preponderance of syllable-onset relative to vowel and coda errors is evidence that long-term constraints on language production are evident in verbal working memory tasks. The first caveat to this interpretation is that on some levels, the distribution of errors we observe does not correspond to those observed in natural speech corpora. Errors in natural speech often exhibit an anticipatory bias, in which the source of the erroneous utterance occurs later in the production plan (Dell, 1986; Shattuck-Hufnagel, 1979). It is clear from looking at Appendix F, however, that in the case of item repetitions, there is no anticipatory bias and perhaps even a slight perseveratory one. Furthermore, when memory demands were imposed in spoken and typed recall, tongue twister effects were observed for noncontextual substitutions, and there were instances in which the syllable-position constraint was violated. We do not think that this result undermines our basic claims, because our error rates are higher than those found in natural speech corpora, and there is evidence that overall task difficulty can affect the distribution of errors. For example, Dell et al. (1997) find a shift away from anticipatory errors toward perseveratory errors as error rates increased, both in their computational model and in their study of several patient groups. Additionally, as production deficits associated with aphasia increase, the extent to which speech error patterns are primarily contextual and abide by the syllable-position constraint decrease (see Dell et al., 1997). Such an effect can be modeled either by increasing the rate with which information decays (Dell et al., 1997), or by degrading the connections between lexical and sublexical levels of representation (Foygel & Dell, 2000).

The second caveat is that the stimuli we used likely exacerbated the syllable-position and syllable-onset effects. Stimuli were designed such that onset speech sounds were very similar, but the codas were less similar. By virtue of this asymmetry alone, one should have expected more onset errors. The same might be said of the syllable-position effects we observed. Vowels are of course unlike consonants, thus they virtually never exchange with consonants. However, the onset and coda consonants we used were also dissimilar to one another, hence their likelihood of substituting may have been minimized. One might argue, then, that the adherence to the syllable-position constraint and the preponderance of onset errors we observed was simply a function of the stimuli we used. In support of this idea, past research has shown that stimuli can be generated in which both of these constraints are violated (Shattuck-Hufnagel, 1992; Taylor & Houghton, 2005). One simple means of testing this possibility would be to design stimuli in which syllable codas are more similar to each than are syllable onsets (e.g. zife pove pife zove), or in which similarity between syllable positions is manipulated (e.g. tith thut thit tuth; though recognizing that at the sub-phonemic level, a phoneme in onset position will not be identical to that phoneme in the coda).

Despite this caveat, we are still confident that the syllable-position and onset effects in these experiments provide evidence that speech production mechanisms underlie our serial ordering errors. First, one might assume that one of the consequences of holding on to the nonwords in this study is decay of information that leads to a break-down of the representations necessary for phonological encoding, hence perfect adherence to these constraints should not have been expected. Second, the vast majority of research into phonological encoding processes has confirmed the existence of these two constraints, both for naturally occurring speech errors (Shattuck-Hufnagel, 1979), and for production of tongue twister stimuli (Sevald & Dell, 1994; Wilshire, 1999), and many computational accounts incorporate these constraints as well (e.g. Dell, 1986; Hartley & Houghton, 1996). Rather than being hard-wired, however, we believe that these constraints emerge as function of learning the distributional properties of the English language, a result that Dell, Juliano, and Govindjee (1993) observed in their computational model of single-word production. Dell et al. argued that the syllable-position constraint emerges from the fact that probability of a vowel following an onset consonant is lower than the probability that codas follow the vowels. In other words, the distribution of speech sounds within the language is such that rhyme units tend to cohere more than onset-vowel combinations (i.e. the CV of a CVC word) do. Moreover, Dell, Reed, Adams, and Meyer (2000) demonstrated that language production processes rapidly learn new statistical regularities concerning the distribution of phonemes in syllables. These results suggest that the syllable-position constraints observed in our studies reflect long-term learning of speech production processes, perhaps augmented by additional learning of the distributional properties of our stimuli, what Dell et al. (2000) call "experiment wide" constraints in production processes. In either case, the existence of

these syllable-position constraints support a language production basis for serial ordering in these studies.

Implications of the language production hypothesis

Since we have managed to complete, and learn from, the extensive analyses presented here, it is probably not surprising that we offer some proselytising about the value of these efforts. Even if the language production hypothesis proves to be incorrect, the use of speech error analyses provides a substantially finer grain of detail to participant performance than item-level accuracy analyses. Such analyses should elucidate present-day theories of verbal working memory or their applications to related fields such as language development and neuropsychology. For instance, regularities in performance across different populations may appear in errors that were once considered noise in recall data (e.g., contextual and noncontextual phoneme substitutions). Other fine-grained analyses besides the ones presented here might also prove to be useful. For example, analyses of self-corrections during serial recall tasks may provide additional insight into the mechanisms responsible for the maintenance and monitoring of serial ordering of phonological information. Beyond the use of error analyses, we would also encourage greater attention to the composition of stimulus lists (essential if long-term learning about the language influences memory performance) and the use of F2 (i.e., items) analyses. This analysis technique necessarily controls for variation across stimulus items. In the present study, we were able to explore the tongue twister effect by comparing stimulus lists composed of the same items presented in different orders.

Beyond the methodological implications, the present investigation demonstrates that serial ordering effects due to phonological similarity reflect sublexical errors in the speech production system (for similar conclusions, see Ellis, 1980; Page et al., 2007). Given this evidence, we feel that the phonological similarity effect can no longer be taken as evidence for language-independent, short-term stores specifically dedicated to maintaining phonological information (e.g. the phonological store; Baddeley, 1986). Quite simply, if the majority of this effect can be explained as sublexical errors within the speech production system, then both descriptive and computational modeling efforts should come to reflect these processes.

We are not the first researchers to make such arguments, as reflected in the fact that computational models of verbal working memory have incorporated aspects of linguistic structure. For example, Hartley and Houghton's (1996) linguistically constrained model short-term memory for nonwords contained two representations of syllable structure: the syllable group layer consisted of two nodes coding syllable onset and rhyme, respectively, and the syllable template layer specified constraints on which phonemes could legally be located within syllable onset a rhyme positions. This same template was suggested for use in a model designed to account for word learning and verbal working memory performance (Gupta & MacWhinney, 1997). Page and Norris (1998a,b) argued that ef-

fects of phonological similarity are explainable as errors in a secondary, speech production-based system, and suggest that this might be implemented by incorporating aspects of Dell's interactive activation model of single-word production (Dell, 1986). Finally, a recent model of nonword repetition developed by Gupta and Tisdale (in preparation) used a simple-recurrent architecture with syllabic representation of stimuli at input and output. The model was designed to address whether nonword repetition ability came from verbal working memory (operationally defined as activation maintenance) or from linguistic knowledge (operationally defined as vocabulary size). Both of these abilities affected the model's performance. However, the critical finding for the present purpose is that even when the model's activation maintenance parameter was held constant, increases in vocabulary size alone could account for developmental increases in nonword repetition ability. In other words, learning the distributional regularities and constraints of sequential ordering in language was sufficient to account for the long-studied correlation between nonword repetition ability and vocabulary acquisition (Edwards, Beckman, & Munson, 2004; Gathercole & Baddeley, 1990).

These models all demonstrate that incorporation of sublexical processes associated with phonological ordering has much to offer in modeling verbal working memory phenomena. We do not want to imply, however, that sublexical phonological encoding is the only level of the production (or the language architecture) over which information is being maintained. Indeed, a justifiable criticism of our current emphasis is that there are numerous errors in working memory tasks that genuinely reflect item-level errors. People often substitute whole items from other lists they have previously encountered (Henson, 1998), and add items that weren't present in any stimulus list. Furthermore, research has shown that the phonological similarity influences both sublexical and list-wise serial ordering (Gupta et al., 2005). In the present study, tongue-twister effects were observed not just for substitutions but also for item omissions in the second and third experiments. Within the language production hypothesis, omission errors would necessarily have to occur at a level that precedes phonological encoding, namely, during the process responsible for lexical-level production planning.

As we suggested before, we see two complementary sources of omission errors. The first is in a failure to properly activate the lexical representation that preceded phonological encoding. In this sense, omission errors reflect a failure to maintain lexical activation (i.e., a "memory" error). Some support for this view comes from the fact that item omissions tended to occur later in the list, as is predicted by some models of verbal working memory (e.g., Page & Norris, 1998a). Even with this account, however, it is unclear why one should expect more omission errors in the tongue twister condition. As such, we feel that omissions may also reflect the functioning of an error monitoring system that prevents people from repeating themselves (see Levelt, Roelofs, & Meyer, 1999). Since participants (presumably) were aware that items did not repeat within a list, we have argued that repetitions likely

reflect a sublexical phoneme error, rather than an error in self-monitoring of entire items. The greater number of omissions in tongue twister conditions may have reflected both self-monitoring for repetition coupled with a failure to activate the correct phonological code. Regardless of the particular account of item omissions, the fact remains that they necessarily reflect an error in maintaining a lexical-level representation, thus sublexical processes alone cannot account for the full range of behaviors.

As it turns out, maintenance of multiple lexical representations is precisely the level over which models of working memory excel (e.g. Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Henson, 1998; Page & Norris, 1998a). To this end, we feel that a recent redefinition of the "phonological loop" as a lexical-level utterance plan (Page et al., 2007) serves as a very useful place from which production researchers might incorporate theoretical and computational ideas from the working memory domain (see Vousden, Brown, & Harley, 2000 for one such example). Some possibilities include developing a model of multi-word utterances in message-driven production by setting up a fast-decaying primacy gradient (Page & Norris, 1998a), and investigating whether mechanism such as competitive queuing (e.g. Houghton, 1990) provides insight into how a grammatical structure or words are chosen from amongst a number of competing alternatives. Production processes must include maintenance of information, as when a message must be maintained while grammatical encoding proceeds, and the output of grammatical encoding must be maintained while driving phonological encoding processes. As such, we suggest that there is much to be learned about planning multi-word utterances in normal production through considering mechanisms that have been proposed in models of verbal working memory.

In sum, our results support Bock's (1996) suggestion that rather than viewing immediate recall in verbal working memory tasks as "the emptying of a short-term store,…the process may be better likened to one of producing a response by assembling highly activated linguistic elements, using the mechanisms of production to do so" (p. 400). The language production hypothesis we have pursued offers a related claim, that the serial ordering processes in verbal working memory tasks can be achieved via the temporary maintenance of representations within the production architecture. In this sense, we view the language production hypothesis as a more explicit version of accounts of working memory as emergent from the temporary activation of long-term perception and action systems (e.g. Cowan, 1995; Postle, 2006). Although we have emphasized sublexical speech errors as the likely source of the phonological similarity effect, we believe that a full account of verbal working memory performance will necessarily take into account levels of production planning that precede (e.g. grammatical encoding and lexical selection) and follow (i.e. articulation) phonological encoding. On this emergentist view, a key question is not what comprises specialized verbal working memory systems, but rather how long-term verbal memory (or linguistic representations) is deployed over short periods of time.

Appendix A. Examples of common speech error types[1]

Substitution
- is a*b*ung our panelists tonight (among).

Exchange
- p*lor*iferation (proliferation).

Shift
- State-*lo*wned _and (owned-land).

Addition
- either the p*l*ubilicity would be bad (publicity).

Omission
- piano sonata _umber ten (number ten).

Appendix B. Nonword stimuli used in Experiments 1 and 2. Tongue twister (TT) and non-tongue twister (Non-TT) orders are presented within a single list.

Tongue twister order	Nonword list position			
	1	2	3	4
TT	Shif	Seev	Sif	Sheev
TT	Tib	Deef	Dib	Teef
TT	Bim	Peeg	Pim	Beeg
TT	Fis	Veem	Vis	Feem
TT	Jiln	Cheal	Chiln	Jeal
TT	Kish	Geam	Gish	Keam
TT	Rish	Leeb	Lish	Reeb
TT	Shife	Sote	Sife	Shote
TT	Fize	Vope	Vize	Fope
TT	Tife	Dorb	Dife	Torb
TT	Jice	Chode	Chice	Jode
TT	Bime	Pise	Pime	Bise
TT	Ribe	Loke	Libe	Roke
TT	Kipe	Gose	Gipe	Kose
Non-TT	Buln	Bape	Puln	Pape
Non-TT	Tup	Tabe	Dup	Dabe
Non-TT	Shul	Shate	Sul	Sate
Non-TT	Fut	Fabe	Vut	Vabe
Non-TT	Ruln	Rabe	Luln	Labe
Non-TT	Kug	Kafe	Gug	Gafe
Non-TT	Jun	Jate	Chun	Chate
Non-TT	Fod	Foose	Vod	Voose
Non-TT	Shar	Shoof	Sar	Soof
Non-TT	Jod	Joof	Chod	Choof
Non-TT	Tof	Toog	Dof	Doog
Non-TT	Koss	Koov	Goss	Goov
Non-TT	Bof	Boov	Pof	Poov
Non-TT	Rart	Roog	Lart	Loog

---

[1] Examples are from Shattuck-Hufnagel (1979). Correct production is provided in parenthesis.

Appendix C. Total number of items with errors (mean proportion and *SD*) for each error type as a function of list position and tongue twister condition in Experiment 1 (rapid reading).

List position	Tongue Twister					Non-tongue twister				
	1	2	3	4	Total	1	2	3	4	Total
Substitutions										
Noncontextual										
Phoneme	7 (0.006, 0.08)	9(0.008,0.09)	22(0.020,0.14)	5(0.01,0.07)	43(0.01,0.1)	1(0.0009,0.03)	8(0.007,0.08)	4(0.04,0.06)	13 (0.01,0.11)	26(0.006,0.076)
Item	0	0	0	0	0	0	0	0	0	0
Contextual										
Phoneme	62(0.06,0.23)	37(0.03,0.18)	68(0.06,0.24)	75(0.07,0.25)	242(0.05,0.23)	24(0.02,0.15)	65(0.06,0.23)	33(0.30,0.17)	31(0.28,0.16)	153(0.03,0.18)
Item	55(0.05,0.22)	36(0.03,0.18)	57(0.05,0.22)	66(0.06,0.24)	214(0.05,0.21)	16(0.01,0.12)	59(0.05,0.22)	23(0.02,0.14)	25(0.02,0.15)	123(0.28,0.16)
Omissions										
Phoneme	4(0.004,0.06)	6(0.03,0.07)	7(0.06,0.08)	0	17(0.004,0.06)	5(0.005,0.07)	1(0.0009,0.03)	6(0.005,0.07)	0	12(0.003,0.005)
Item	7(0.006,0.08)	6(0.005,0.07)	7(0.006,0.08)	4(0.004,0.06)	24(0.005,0.07)	1(0.0009,0.03)	1(0.0009,0.03)	1(0.0009,0.03)	0	3(0.0007,0.03)
Additions										
Phoneme	2(0.002,0.04)	2(0.002,0.04)	3(0.003,0.05)	0	7(0.002,0.04)	2(0.002,0.04)	0	4(0.004,0.06)	0	6
Item	0	0	0	0	0	0	0	0	0	0

Appendix D. Total number of items with errors (mean proportion and *SD*) for each error type as a function of list position and tongue-twister condition in Experiment 2 (spoken recall).

List position	Tongue twister					Non-tongue twister				
	1	2	3	4	Total	1	2	3	4	Total
Substitutions										
Noncontextual										
Phoneme	19(0.055,0.23)	9(0.26,0.16)	24(0.07,0.25)	16(0.05,0.21)	68(0.05,0.22)	15(0.04,0.20)	6(0.02,0.13)	20(0.05,0.23)	5(0.01,0.11)	46(0.03,0.17)
Item	0	0	0	0	0	0	0	0	0	0
Contextual										
Phoneme	71(0.20,0.40)	115(0.33,0.47)	110(0.32,0.47)	106(0.30,0.46)	402(0.29,0.45)	59(0.16,0.37)	73(0.20,0.40)	44(0.12,0.33)	44(0.12,0.33)	220(0.15,0.36)
Item	63(0.18,0.39)	105(0.30,0.46)	99(0.28,0.45)	95(0.27,0.45)	362(0.26,0.44)	55(0.15,0.36)	63(0.17,0.38)	32(0.09,0.28)	39(0.11,0.31)	189(0.13,0.34)
Omissions										
Phoneme	10(0.03,0.17)	0	9(0.03,0.16)	1(0.001,0.05)	20(0.01,0.12)	8(0.02,0.15)	2(0.01,0.07)	9(0.020.16)	0	19(0.01,0.11)
Item	15(0.04,0.20)	12(0.03,0.18)	46(0.13,0.34)	56(0.16,0.37)	129(0.09,0.29)	4(0.01,0.10)	4(0.01,0.10)	17(0.05,0.21)	21(0.06,0.23)	46(0.03,0.17)
Additions										
Phoneme	1(0.001,0.05)	6(0.02,0.13)	2(0.01,0.08)	4(0.01,0.11)	13(0.01,0.10)	1(0.001,0.05)	1(0.001,0.05)	2(0.003,0.07)	0	4(0.001,0.05)
Item	0	0	0	0	0	0	0	0	0	0

Appendix E. Total number of items with errors (mean proportion and *SD*) for each error type as a function of list position and tongue twister condition in Experiment 3 (typed recall).

List position	Tongue twister					Non-tongue twister				
	1	2	3	4	Total	1	2	3	4	Total
Substitutions										
Noncontextual										
Phoneme	31(0.13,0.33)	26(0.11,0.31)	36(0.15,0.35)	29(0.12,0.32)	122(0.12,0.33)	17(0.68,0.25)	23(0.09,0.29)	20(0.79,0.27)	19(0.08,0.26)	79(0.08,0.27)
Item	0	0	0	0	0	0	0	0	0	0
Contextual										
Phoneme	52(0.21,0.41)	60(0.24,0.43)	62(0.25,0.43)	93(0.38,0.49)	267(0.27,0.44)	10(0.04,0.20)	45(0.18,0.38)	34(0.13,0.34)	41(0.16,0.37)	130(0.13,0.34)
Item	35(0.14,0.35)	46(0.19,0.39)	43(0.17,0.38)	77(0.31,0.46)	201(0.20,0.40)	5(0.02,0.14)	24(0.10,0.29)	17(0.07,0.25)	18(0.07,0.26)	64(0.06,0.24)
Omissions										
Phoneme	4(0.02,0.13)	1(0.004,0.063)	9(0.04,0.19)	2(0.008,0.09)	16(0.02,0.13)	3(0.01,0.11)	2(0.008,0.089)	2(0.008,0.089)	5(0.02,0.14)	12(0.01,0.11)
Item	13(0.05,0.22)	13(0.05,0.22)	19(0.07,0.27)	18(0.07,0.26)	63(0.06,0.24)	1(0.004,0.06)	2(0.008,0.09)	7(0.028,0.16)	8(0.03,0.18)	18(0.02,0.13)
Additions										
Phoneme	1(0.004,0.06)	2(0.008,0.09)	5(0.02,0.14)	3(0.012,0.11)	11(0.011,0.11)	1(0.004,0.06)	2(0.008,0.09)	8(0.03,0.18)	2(0.008,0.09)	13(0.01,0.11)
Item	0	0	0	0	0	0	0	0	0	0

Appendix F. Breakdown of item contextual substitutions (mean proportion) into exchanges and repetitions across tongue twister conditions and proportion of anticipation and perseveration errors for repetitions across Experiments 1–3.

	Tongue twister				Non-tongue twister		
Experiment 1: rapid reading							
Exchanges	36 (0.17)				36 (0.29)		
Repetitions	158 (0.74)	Anticipations	48 (0.30)	49 (0.40)	Anticipations	13 (0.27)	
		Perseverations	110 (0.70)		Perseverations	36 (0.73)	
Experiment 2: spoken recall							
Exchanges	229 (0.63)				136 (0.72)		
Repetitions	128 (0.35)	Anticipations	40 (0.31)	53 (0.28)	Anticipations	27 (0.51)	
		Perseverations	88 (0.69)		Perseverations	26 (0.49)	
Experiment 3: typed recall							
Exchanges	120 (0.57)				36 (0.55)		
Repetitions	90 (0.43)	Anticipations	30 (0.33)	29 (0.45)	Anticipations	10 (0.34)	
		Perseverations	60 (0.67)		Perseverations	19 (0.66)	

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and phonological encoding in speech production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135*, 50–68.

Allport, A. D. (1984). *Auditory-verbal short-term memory and conduction aphasia*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Baddeley, A. D. (1986). *Working memory*. New York, NY, US: Clarendon Press/Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 36*, 233–252.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589.

Bock, J. K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3*, 395–421.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127–181.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551–581.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior, 15*, 17–31.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75–84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429–432.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY, US: Oxford University Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 222–233.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283–321.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104*, 123–147.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science, 17*, 149–195.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Cognition, 26*, 1355–1367.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801–838.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language and Hearing Research, 47*, 421–436.

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior, 19*, 624–634.

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

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43*, 182–216.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27–52.

Garrett, M. F. (1975). *The analysis of sentence production*. San Diego, CA: Academic Press (pp. 133–175).

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200–213.

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*, 439–454.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 84–95.

Gupta, P., & Tisdale, J. (in preparation). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate.

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition, 33*, 1001–1016.

Gupta, P., & MacWhinney, B. (1995). Is the articulatory loop articulatory or auditory? Reexamining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language, 34*, 63–88.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59*, 267–333.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35*, 1–31.

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*, 73–137.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). London: Academic Press.

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685–701.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences, 10*, 480–486.

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 656–674.

Levelt, W. J. M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

Levy, B. (1971). Role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behaviour, 10*, 123–132.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*, 323–350.

Meijer, P. J. A. (1997). What speech errors can tell us about word-form generation: The roles of constraint and opportunity. *Journal of Psycholinguistic Research, 26*, 141–158.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78*, 679–684.

Nimmo, L. M., & Roodenrys, S. (2004). Investigation the phonological similarity effect: Syllable structure and the position of common phonemes. *Journal of Memory and Language, 50*, 245–258.

Nooteboom, S. G. (1969). The tongue slips into patterns. In A. G. Sciarone, A. J. van Essen, & A. A. Van Raad (Eds.), *Leyden studies in linguistics and phonetics* (pp. 114–132). Mouton: The Hague.

Page, M. P. A., Madge, A., Cumming, N., & Norris, D. G. (2007). Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. *Journal of Memory and Language, 56*, 49–64.

Page, M. P. A., & Norris, D. (1998a). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761–781.

Page, M. P. A., & Norris, D. (1998b). Modeling immediate serial recall with a localist implementation of the primacy model. In J. Grainger (Ed.), *Localist connectionist approaches to human cognition* (pp. 227–255). Mahwah, NJ: Lawrence Erlbaum Associates.

Postle, B. R. (2006). Working memory as an emergent property of mind and brain. *Neuroimage, 30*, 950–962.

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 29–33.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 1019–1034.

Saito, S., & Baddeley, A. D. (2004). Irrelevant sound disrupts speech production: Exploring the relationship between short-term memory and experimentally induced slips of the tongue. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57*, 1309–1340.

Salame, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 150–164.

Sevald, C., & Dell, G. (1994). The sequential cueing effect in speech production. *Cognition, 53*, 91–127.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W. E. Cooper & E. C. Walker (Eds.), *Sentence processing: psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Lawrence Erlbaum.

Shattuck-Hufnagel, S., Keller, E., & Gopnik, M. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In *Motor and sensory processes of language* (pp. 17–51). Lawrence Erlbaum Associates, Inc.

Shattuck-Hufnagel, S. (1994). The role of word structure in segmental serial ordering. *Cognition, 42*, 213–259.

Taylor, C., & Houghton, G. (2005). Learning artificial phonotactic constraints: Time course, durability and relationship to natural constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*(6), 1393–1416.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 145–152.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41*, 101–175.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 1256–1271.

Wickelgren, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology, 70*, 102–108.

Wilshire, C. E. (1999). The 'tongue twister' paradigm as a technique for studying phonological encoding. *Language and Speech, 42*, 57–82.