

The distribution of speech errors in multi-word prosodic units

Wook Kyung Choe & Melissa A. Redford

The University of Oregon

Running title:

Speech errors within multi-word prosodic units

Please address correspondence to:

Melissa A. Redford

Linguistics Department

1290 University of Oregon

Eugene, OR 97403

redford@uoregon.edu

ABSTRACT

Sequencing errors in natural and elicited speech have long been used to inform models of phonological encoding and to understand the process by which serial ordering is achieved in speech. The present study focused on the distribution of sequential speech errors within multi-word prosodic units to determine whether such units are relevant to speech planning, and, if so, how. Forty native English-speaking undergraduate students were asked to produce sentences that varied in length and in the extent to which certain phonological features were repeated (tongue twisters or not). Participants prepared their utterances in advance of speaking and were coached to be as fluent as possible once they started speaking. The goal was to ensure the production of well-structured utterances, while maximizing the number of errors produced, and minimizing the effects that excessive self-correction might have on prosodic structure. Speech errors were perceptually identified in the recorded speech and categorized. Strong and weak prosodic boundaries were prosodically-transcribed in sentences with sequencing errors. Speech error patterns were found to correspond well with the boundaries of the multi-word prosodic units defined by the strong and weak prosodic boundaries. In particular, the number of sequencing errors was found to vary as a function of position within a unit such that the fewest errors were found in initial position, more occurred in early-mid position, and even more occurred in late-mid position. This pattern of increasing errors across the multi-word prosodic unit was referred to as the cumulative error pattern. The analyses also revealed a final position effect. When multi-word prosodic units occurred in utterance-initial or utterance-medial position, a disproportionate number of errors occurred in the final position of the unit. However, when the units occurred in utterance-final position, more errors occurred in late-mid position than in final position. The cumulative error pattern and final position effect are interpreted to suggest the serial activation and decay in activation of multi-word planning domains during phonological encoding.

1. Introduction

In her classic paper, Fromkin (1971) argued for the reality of phonological units such as features, phonemes, and syllables on the basis of erroneous transpositions or displacements of these units in natural speech. She argued, for instance, that syllables were units of representation because when features, phonemes, or clusters were transposed or displaced from a target, they usually preserved the target's syllable position (e.g., *carp-si-hord* for *harp-si-chord*; *ibid.*:p. 39). Since Fromkin's paper, many researchers have used the same sequencing errors not only to understand the psychological reality of linguistic units, but also to understand the architecture of the speech plan and serial ordering processes in speech. For example, Shattuck-Hufnagel (1979, 1983) proposed a scan-copier model to account for the distribution of different types of sequencing errors found in the MIT corpus of spontaneous speech errors and in elicited speech error data. She noted that most of the phonological errors transposed, anticipated, or perseverated syllable onsets, syllabic nuclei, and codas. She also noted that these types of speech errors typically involved a single segment or cluster and were rarely better described as involving distinctive features, rhymes or whole syllables. To account for these distributional facts, Shattuck-Hufnagel proposed that a structurally tagged corpus of candidate segments is deposited into a memory buffer during phonological encoding. This corpus is then scanned and copied by a serial order processor into stressed and unstressed syllabic frames. The suprasegmental (syllable or foot) frames are themselves readied and serially ordered according to the specifications of the lexical items involved, which are ordered prior to phonological encoding according to a syntactic processing component.

Psycholinguistic models of speech production have incorporated many of Shattuck-Hufnagel's (1979, 1983) insights, but the focus has shifted from the units and structure of the speech plan to the storage and sequential retrieval of elements. These models posit that the distributed representation and gradient activation of linguistic elements during speech planning can result in speech errors [see Goldrick (2007) for a recent review of these models]. For

Speech errors in multi-word prosodic units

example, in Dell's (1986) influential model of speech production, morpho-lexical and phonological material are selected via spreading activation following decision rules generated at a higher level of representation. As in Shattuck-Hufnagel's model, selected items are then slotted into appropriate positions within phrasal and word frames that are generated by syntactic and morphological rules and by the phonological entries associated with different morphemes. Crucially, though, the morphemic and phonological items for selection are organized according to similarity and networked so that their selection via activation entails the concomitant activation of multiple, similar items. Although the item with the highest activation level guides the next level of speech planning (lexical → phonological → articulatory), the gradient activation of linked nodes in a network allows for some error in the selection process. As in natural language, model generated errors are uncommon because the most highly activated node in the network of related items is usually the target item, which gets a boost in activation when selected while related items are inhibited. Activation also decays fairly rapidly after selection so that the same item is not mistakenly reselected later.

Apart from explaining the production and distribution of speech errors in psychologically plausible terms, network models emphasize the dynamic nature of speech production: the models account for effects of both the future and the past on the present. With respect to the dynamics, speech planning is treated as an incremental process that occurs at multiple levels of organization simultaneously. Although lower levels in the speech plan (i.e., plans that are more local and detailed with respect to execution) follow from higher-level plans, selection and execution of materials may begin before the plans at higher levels are fully formulated (Stemberger 1985; MacKay 1987; Dell et al. 1997; Goldrick 2007). Even though proper sequencing of material for execution is maintained by a system of activation, inhibition, and decay that is intrinsic to each level in the plan, the particular state of the system at a higher level will have consequences for processing at a lower level. Empirical support for this idea comes from errors that appear to have a mixed relationship to the target (i.e., errors that are

both semantic and phonological or both phonological and articulatory). To take one example, Goldrick and Blumstein (2006) document the effects of current and subsequent phonological targets on the phonetic realization of stop consonant voicing in a tongue twister task designed to elicit speech errors. Their phonetic data show that anticipatory and perseveratory errors have traces of the phonological target: the voice onset values of stops transcribed as voiced were longer when the target was a voiceless stop than when it was a voiced stop and vice versa for the stops that were transcribed as voiceless. Goldrick and Blumstein argued that this type of mixed error indicated that the state of the system at the phonological level influenced articulation via cascading activation even before the whole articulatory action had been completed. In essence, Goldrick and Blumstein suggest that processes at one level of representation can affect activation at another level of representation. This suggestion is distinct from that which is implicit in the architecture of other network models where activation levels can spread only within a single level of representation (i.e., across nodes that encode the same kind of unit).

In spite of their emphasis on the psychologically plausible, the elemental nodes in network models are typically phonemic and syllabic units, and the higher-level nodes are typically syntactically- and semantically-tagged lexemes. The result of this architecture is that the models provide robust accounts of word-level production phenomena, but do not provide much insight into connected speech (but see Levelt 1989, and Levelt et al. 1999). At the same time, the focus in linguistics has moved beyond word-level sound patterns and onto phrase-level sound patterns. In particular, phrase-level accents, once treated primarily in terms of their communicative function (Crystal 1969; Bolinger 1989), are now situated in theories of hierarchical prosodic structure (Beckman and Pierrehumbert 1986; Nespor and Vogel 1986; Selkirk 1995; Shattuck-Hufnagel and Turk 1996; Ladd 2000; Gussenhoven 2002). Disagreements exist regarding the number and exact nature of the units in this hierarchy, but most theories posit a minimum of two high-level units. The larger of these is called the Intonational Phrase

and, though often coterminous with an utterance, is distinct from utterances in that these can be made up of more than one Intonational Phrase. The smaller high-level prosodic unit has been alternately labeled as the Intermediate Phrase (Beckman and Pierrehumbert 1986), the Phonological Phrase (Nespor and Vogel 1986), or the Major Phrase (Selkirk 1995), and it is nested within the Intonational Phrase. Like the larger prosodic unit, the smaller unit often corresponds to clauses or other meaningful chunks of speech. But both units are principally recognized by their intonational and rhythmic coherence and by their boundary characteristics, which include changes in pitch and duration. While the boundary characteristics associated with Intonational Phrases are highly salient, those associated with the smaller unit are somewhat less salient, leading perhaps to the disagreement between authors as to their exact definition. We will gloss over this disagreement by prioritizing the perceptual coherence of multi-word prosodic units and we will recognize the similarity between smaller and larger units by referring to both as Intonational Units (IUs).

Although it is not completely clear how IUs are represented in the grammar, there is ample evidence to suggest that these units represent important speech processing domains. IU boundaries appear to influence language comprehension (see Cutler et al. 1997; Frazier et al. 2006), pause duration (Ferriera 1993; Krivokapic 2007), and speech articulation (Fougeron and Keating 1997; Byrd and Saltzman 2003; Tabain 2003; Cho and McQueen 2005). For example, Fougeron and Keating (1997) investigated the articulation of consonants and vowels at the beginning, middle, and end of larger and smaller prosodic units using electropalatography. The results showed that unit-initial segments were articulated with more linguopalatal contact than unit-medial or -final segments, and that this positional effect was stronger for larger prosodic units (e.g., Intonational Phrases) than for smaller ones (e.g., Prosodic Words). Fougeron and Keating labeled this effect *initial strengthening*, suggesting that the changes may be associated with increased effort or energy at the beginning of an IU.

Croot, Au, and Harper (2010) investigated the effects of phrase position and prosodic prominence on speech errors. They found that errors occurred less often in prominent words generally and were especially infrequent in prominent words that occurred in phrase-initial position. Croot and colleagues interpreted this finding to support Keating and Shattuck-Hufnagel's (2002) view that higher levels of prosodic structure are specified before word forms are encoded in phonological and phonetic detail (contra Levelt 1989, and Levelt et al. 1999). Like Keating and Shattuck-Hufnagel (2002), Croot and colleagues referenced a scan-copier model to discuss serial ordering within multi-word prosodic frames. They argued that elements must be tagged for prominence early on. Such tagging would reduce the number of errors associated with prosodically prominent words because few words would be so tagged in any given utterance. The positional effect was similarly explained. Like Fougeron and Keating (1997), Croot et al. posited that initial position has a unique status. However, Croot et al. were not certain whether this unique status was associated with phrase-initial strengthening or with the study design. Croot et al. pointed out that the two types of tongue twisters they used in their study had the same initial sequences, but different medial and final ones (ABAB vs. ABBA tongue twisters).

Assuming that the positional effect on errors that Croot et al. (2010) reported is a prosodic effect, then we might expect to see the same effect at IU boundaries in utterance-medial position. The present study tested this expectation by looking at the distribution of errors across IUs in different sentence positions. An additional goal was to determine whether prosodically-related positional effects are best interpreted within a scan-copier model or a network model. If elements destined for initial position are uniquely tagged and other positions are not, then the protection afforded by tagging should not extend to other positions and errors should be equally high across all non-initial positions within the unit. However, if a positional effect on errors is related to the activation of an IU-sized domain for phonological encoding, and errors are minimized when activation levels are high, then we might posit an

asymmetric distribution of errors across the length of an IU that would be consistent with high initial activation and subsequent decay in activation.

2. Methods

2.1. *Participants*

40 native English-speaking University of Oregon undergraduates participated in the study, 19 participants were male and 21 were female. All participants received course credit for their participation.

2.2. *Stimuli*

The stimuli were 60 different sentences that varied in length such that there were 20 short sentences (4 to 7 syllables; $M = 4.50$ words), 20 medium sentences (8 to 14 syllables; $M = 6.65$ words), and 20 long sentences (15 to 25 syllables; $M = 14.10$ words). The length of sentences was manipulated to encourage the natural prosodification of the stimuli into different sized IUs.

The sentences also varied in the extent to which phonological features were repeated: 30 sentences were tongue twisters, and 30 were not. Tongue twisters were used in order to induce sufficient errors for a meaningful distributional analysis. The average error rate in spontaneous speech is extremely low: Levelt (1989) suggests, for instance, that it is less than 0.1%. Non-tongue twisters were included so that we could at least qualitatively assess whether the error patterns elicited using tongue twisters might be similar to those which occur in more typical speech.

The large number of different sentences that we used also provided extensive phonological, semantic, and syntactic variability. Such variability is akin to the variability that is introduced into experiments by randomly sampling from a larger population.

Participants repeated 60 different sentences 5 times. Specifically, a total of 300 sentences were presented to participants in one of 3 pre-determined randomized orders.

2.3. *Procedure*

A single participant was seated in a sound-attenuated booth in front of a computer monitor and a standing microphone. The stimulus sentences were presented on the monitor, first in red and then in green. The participant was asked to prepare the entire sentence by reading it silently when presented in red on the screen. Once the participant felt prepared to speak, s/he was told to press a button at which point a beep sounded and the entire sentence turned from red to green, signaling that the participant should start speaking. The purpose of this procedure was to maximize the likelihood that participants would construct a well-structured plan to guide speech output rather than reading the sentences word for word. Preparation times suggested that participants did in fact plan their output: a repeated measures ANOVA on the average preparation times for each speaker indicated that length significantly affected planning times [$F(2,78) = 58.59, p < .01$]; post-hoc mean comparisons showed that speakers planned long sentences for longer than medium length sentences (mean difference = 603.26, $SE = 81.12, p < .01$), which they planned for longer than short ones (mean difference = 395.96, $SE = 53.22, p < .01$).

Participants were instructed to speak at a comfortable rate. They were also instructed not to stop, pause, or self-correct while they were speaking. The goal of these instructions was to maximize the naturalness of speakers' prosody while also maximizing speech fluency as well as the number of errors produced. An experimenter remained with the participant throughout the session to provide feedback about speech rate and speaking level as well as to remind participants to speak as fluently as possible regardless of what errors they made.

Participants were given 10 trials to become familiar with the procedure before the test stimuli were presented. The entire session was digitally recorded for later perceptual analysis.

2.4. Coding

2.4.1. Speech errors

The first author (WKC) and an undergraduate research assistant listened to the 12,000 sentences produced by the 40 participants and independently identified sentences that contained speech errors. Mispronunciations and disfluencies were also noted, but sentences that contained only these were not considered for further analysis. Inter-judge agreement was 93.94%. Disagreements were settled according to the opinion of the second author (MAR), who listened only to the 1,397 sentences that were identified as containing one or more speech errors.

Next, the authors independently categorized speech errors as errors of anticipation, perseveration, anticipation and perseveration, lexical substitution, insertion, or deletion. The error category depended on the presence and location of a source for the error within the stimulus sentence or preceding 2 or 3 stimulus sentences. Errors were categorized as anticipatory when the closest possible source for the error was subsequent to the error (e.g., “Greep grapes are great” instead of “Greek grapes are great”). Errors were categorized as perseveratory when the closest possible source for the error preceded the error (e.g., “Fred frew three free throws” instead of “Fred threw three free throws”). Errors were categorized as both anticipatory and perseveratory when the subsequent and preceding material could have provided the source and both were equally close to the error (e.g., “Friendly Frank flips fline flapjacks” instead of “Friendly Frank flips fine flapjacks”). Errors were categorized as errors of lexical substitution (e.g., “One smart fellow, he was smart” for “One smart fellow, he felt smart”), deletion (e.g., “Three grey geese in the ___ grass grazing” for “Three grey geese in the green grass grazing”), or insertion (e.g., “Glenwood makes fine apple flapjacks” for “Glenwood makes ___ apple flapjacks”) depending on the type of error and when no possible source for the error could be identified within the sentence or from preceding set of sentences. The

authors were in agreement on 94.72% of the categorizations. Disagreements were resolved through repeated listening and discussion.

As indicated above, only speech errors with a clear source were categorized as errors of anticipation, perseveration, or both and the nearest possible source was always chosen, even when the source could have come from further away. For example, we would identify the immediately preceding word “black” as the source of the error “blear” in the sentence “*A big black bug bit a big black blear, making the big black bear bleed blood*” even though there are other candidate source words for the error in this sentence. Even with the criterion of nearest source, we found that the source and error were often separated by more than one word. Accordingly, we coded the relationship between the source and the error in one of four ways: (i) within word, (ii) adjacent word, (iii) non-adjacent word within the sentence, and (iv) word from a preceding sentence.

2.4.2. *Boundaries*

After all of the speech errors had been identified, the authors identified prosodic boundaries. Boundary identification was originally limited to those sentences with errors of anticipation, perseveration, and both anticipation and perseveration ($N = 966$).

The boundary identification procedure was honed on the data from 16 participants. In addition to coding the location of boundaries, we also decided to code 2 levels of prosodic boundary strength: strong and weak. We repeatedly checked our perception of boundary location and strength against spectrograms and F0 traces of the sentences. These spot checks confirmed our impression that strong boundaries were associated with final lengthening, a boundary tone, and a pause. Weak boundaries were mostly identified when a pause was absent, but a boundary tone and some degree of final lengthening was present. Speech disfluencies due to self-correction or hesitation were coded separately and so are not confounded with the boundaries that are the focus of this report.

Whereas we jointly located strong and weak boundaries for 16 speakers, we independently identified these in the other 24 participants' data. Inter-author agreement on location and strength for these data was 94.43%. Disagreements were resolved through repeated listening and by checking the visual representations of the sentences.

Once boundaries were identified in the sentences with speech errors, one of us (WKC) used our agreed-upon criteria to identify prosodic boundaries in a matched sample of sentences without speech errors ($N = 966$). The goal of this exercise was to determine whether or not the prosodic structures of sentences with speech errors was similar to sentences without speech errors. Sentences without speech errors were selected speaker by speaker to match in length and phonological repetitiveness the sentences that individual speakers produced with speech errors. These sentences were then transcribed for boundary location and strength. To ensure transcription reliability, the second author independently coded boundary location and strength in a randomly selected subsample of the sentences without errors ($N = 250$). Inter-author reliability was 95.20%.

3. Results and Discussion

Speakers produced a total of 1,634 speech errors. Errors of anticipation and/or perseveration were the most frequent types ($N = 1,257$). The fewest number of these errors occurred in IU-initial position, even when the IU boundary was in sentence-medial position. Although more errors occurred in subsequent IU positions, a disproportionate number occurred next to a final strong or weak IU boundary, especially in sentence-nonfinal position. IU-final errors in sentence-nonfinal position were equally likely to have their source from elements that were on the other side of an IU boundary (i.e., errors of anticipation) as they were to have their source from the same side of the boundary (i.e., errors of perseveration).

3.1. *Speech Errors*

On average, participants produced speech errors in 8.07% of the stimulus sentences, though error rates ranged from 1% to 17.33% across participants, the standard deviation in error rate was 4.04%. Altogether a total of 1,634 speech errors were identified in 1,397 sentences, the majority of which were tongue twisters (77.52%). Of the 1,634 errors, 681 were identified as errors of anticipation, 507 were identified as errors of perseveration, 69 were identified as errors of both anticipation and perseveration, and 377 were identified as other types of errors (i.e., lexical substitution, deletion, or insertion). Since, by definition, only errors of anticipation and/or perseveration are sequencing errors, and since these constituted the majority of all errors, the remaining analyses will focus on these types of errors.

The number of anticipatory and/or perseveratory errors that participants produced varied with sentence length. The fewest such errors were produced in short sentences ($N = 205$), more were produced in medium length sentences ($N = 463$), and the most were produced in long sentences ($N = 589$). Next, we turn to the results from the independent coding of prosodic boundaries in the 966 sentences that contained the 1,257 errors of anticipation and/or perseveration and those in the matched set of 966 sentences that contained no errors.

3.2. *Prosodic Boundaries*

Table 1 shows the number of sentences with strong and weak IU boundaries as a function of sentence length and the existence of speech errors in the sentences. Not surprisingly, a greater percentage of long sentences compared to medium or short sentences had at least one sentence-internal strong or weak boundary (77.11% in long sentences versus 31.52% in medium length and 21.16% in short sentences). The table also indicates that the distribution of boundaries in sentences with and without errors was similar across all sentence lengths and boundary types. However, there were more internal IU boundaries in sentences

with speech errors than in those without speech errors. This difference was significant in a chi-square test ($\chi^2 (1, N = 1932) = 46.61, p < .01$).

Table 1. Sentences with different types of internal boundaries are shown as a function of sentence length and whether the sentence was produced with or without speech errors. The total number of sentences represented in the table sums to more than 1,932 (966 x 2) because a single sentence could be produced with both a strong and weak sentence-internal IU boundary.

Sentence type	Sentence length	No internal boundary	1 or more strong boundary	1 or more weak boundary
With Error	Short	115	2	46
	Medium	214	7	130
	Long	67	97	362
	TOTAL	396	106	538
No Error	Short	143	2	22
	Medium	264	3	83
	Long	139	70	295
	TOTAL	546	75	400

The mean length of units defined by IU boundaries also varied with sentence length. Of course, short sentences with no internal boundaries were shorter than medium sentences with no internal boundaries ($M = 4.65$ words, $SD = 0.74$ vs. $M = 6.96$ words, $SD = 1.24$), which were in turn shorter than long sentences with no internal boundaries ($M = 6.96$ words, $SD = 1.24$ vs. $M = 11.18$ words, $SD = 3.11$). This pattern also held for the sentence-internal units defined by either strong or weak boundaries. Such units were shortest in short sentences ($M = 2.54$ words, $SD = 1.06$), longer in medium sentences ($M = 3.08$ words, $SD = 1.42$), and longest in long

sentences ($M = 5.15$ words, $SD = 2.37$). Mann-Whitney U tests comparing these unit lengths in sentences with and without errors indicated that these were only significantly different in medium length sentences where unit lengths were an average of 0.52 words shorter in the sentences with errors than in those without errors ($z = -4.1$, $p < .01$).

3.3. *Speech Errors and Prosodic Boundaries*

Scatterplots were used to evaluate the distribution of errors across positions in units of varying lengths. Position was defined as the position of the word within a unit that contained a speech error, counting from the beginning of a unit. Unit length was defined by number of words. Three types of units were considered: utterances, big IUs, and small IUs. Utterances were defined by a single prosodic boundary, namely, the boundary that coincided with the sentence boundary. Big IUs were defined by a sentence-internal strong boundary and either another sentence-internal strong boundary or the beginning or ending of the sentence. A small IU was defined by a sentence-internal weak boundary on one side and either a strong or weak boundary on the other side. Note that this definition means that big IU and small IU refer only to boundary strength, not to the size or length of an IU.

Regression line slopes were plotted for each scatterplot. These lines indicated a positive relationship between the position of an error within a unit and the number of words within that unit, as shown in Figure 1. This relationship was stronger for big and small IUs than for utterances [error-position/word-number line: utterance, slope = .52, $R^2 = .31$; big IU, slope = .81, $R^2 = .48$; small IU, slope = .77, $R^2 = .70$].

Speech errors in multi-word prosodic units

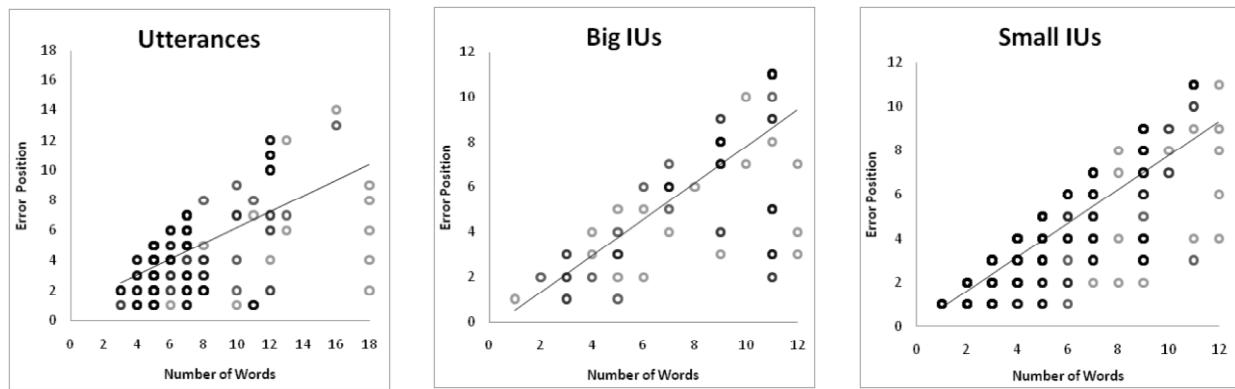


Figure 1. The position of words containing speech errors (error position) is shown as a function of the number of words in utterances, big IUs, and small IUs. The circles in the figure represent errors. Darker circles correspond to greater numbers of errors for that particular position and unit length than lighter circles.

If errors had been distributed randomly across a unit, then the slope of the error-position by word-number regression line would have been .5. Instead, the steeper slopes indicate that the fewest errors occurred in initial position and the greatest number of errors occurred towards the end of a unit. This positional effect was much stronger in big and small IUs than in utterances, where the slope was only barely above .5. Then again, the shallower slope in utterances was likely due to the especially wide distribution of errors in the longest units (see Figure 1). When the slope was recalculated for utterances with 12 or fewer words, which was the maximum number of words in big and small IUs, it was much steeper (slope = .66, $R^2 = .36$).

The especially steep slopes of the error-position by word-number regression line for big and small IUs and the high proportion of variance accounted for by these lines suggest that the positional effect was related to the IU rather than to the sentence as a whole. To confirm that this was indeed the case, we divided the data into units that occurred early or in the middle of a sentence (sentence-nonfinal IUs) and those that occurred at the end of a sentence (sentence-final IUs), and then reanalyzed the effect of number of words within a unit on position of error

Speech errors in multi-word prosodic units

within that unit. The data from big and small IUs were combined, since the previous analyses had indicated that the positional effect was similar for both types of units.

The scatterplots for sentence-nonfinal ($N = 447$) and sentence-final IUs ($N = 298$) are shown in Figure 2. In both cases, fewer errors occurred at the beginning of the unit than towards the end of the unit: the error-position by word-number regression line had a slope of .82 ($R^2 = .71$) for sentence-nonfinal IUs and a slope of .77 ($R^2 = .74$) for sentence-final IUs. The fact that the positional effect was equally strong for sentence-nonfinal and sentence-final IUs confirms the suggestion that the distribution of errors is related to the IU and not to sentence boundaries.

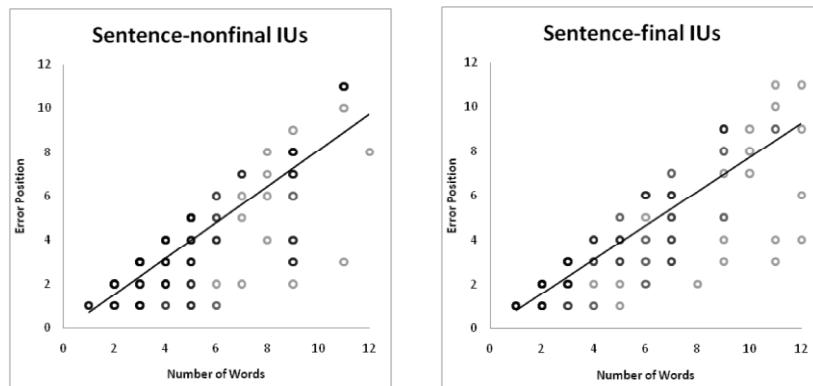


Figure 2. The position of words containing speech errors (error position) is shown as a function of the number of words in sentence-nonfinal and sentence-final IUs. The circles in the figure represent individual errors. Darker circles correspond to greater numbers of errors for that particular position and unit length than lighter circles.

Our next analysis was aimed at answering the question of whether the positional effect was due to a gradual increase in errors across the unit or to a disproportionate number of errors in unit-final position compared to unit-initial position. Accordingly, we compared the number of speech errors that occurred in initial, early, late, and final position within sentence-nonfinal and sentence-final IUs of 4 or more words. Initial position was defined as the first word in the IU; early position as words in the first half of the IU, starting with the second word; late

Speech errors in multi-word prosodic units

position as the second half of the IU, minus the last word in the IU; and, final position as the last word in the IU. A finer grain count was also made of just IUs with odd numbers of words (5, 7, etc.). This count investigated the distribution of speech errors across 5 position categories, including a "middle" category. The patterns were found to exactly parallel those based on the previously described 4 position categories: initial, early, late, final. For this reason, only the results from the overall count are reported below.

The counts by position, shown in Figure 3, indicated that errors were cumulative. That is, errors increased with position number in a unit compared with earlier in that unit.

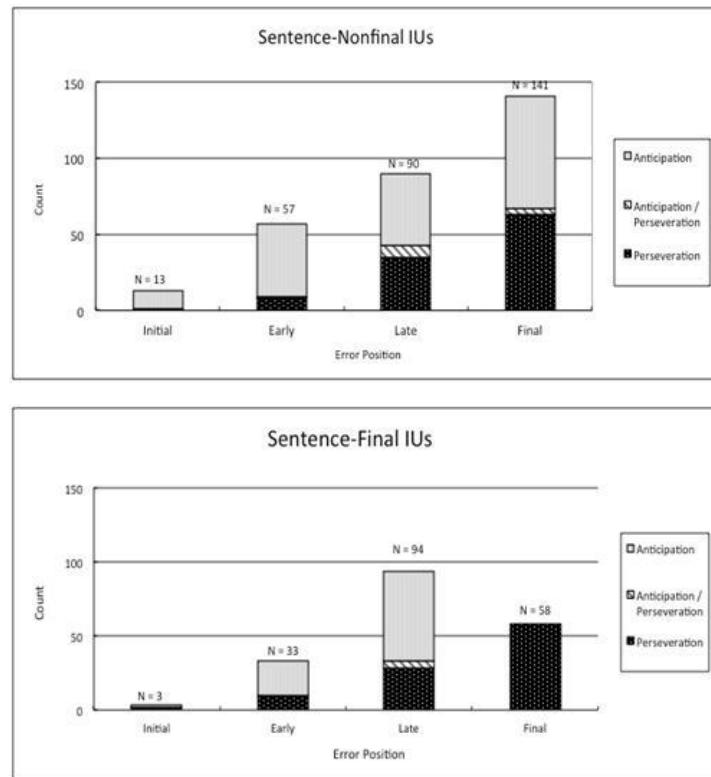


Figure 3. The number of errors in medium and long IUs (4+ words) is shown as a function of the position of the words containing speech errors (error position). The proportion of errors of anticipation and/or perseveration is shown for each position as well. The top panel shows error counts for sentence-nonfinal IUs and the bottom panel shows those for sentence-final IUs.

The cumulative pattern of errors also held for utterances: there were 38 errors in initial position, 121 in early position, 196 in late position, and 145 in final position. That is, the majority of errors (68.2%) occurred in the latter half of the utterance. To investigate whether the pattern also held for units of less than 4 words, we compared the number of unit-initial and unit-final errors in sentence-nonfinal and sentence-final IUs with 2 and 3 words. The results indicated that even in these smaller units, errors were much more likely to occur towards the end of the unit than the beginning (sentence-nonfinal IUs, initial vs. final position, $N = 28$ vs. $N = 82$; sentence-final IUs, initial vs. final position, $N = 21$ vs. $N = 40$).

Figure 3 also shows that the position of peak errors differed depending on whether the IU occurred within or at the end of a sentence. Specifically, most errors occurred in final position for sentence-nonfinal IUs, but in late position for sentence-final IUs and utterances. This difference suggests that the large number of unit-final errors in sentence-nonfinal IUs may be due to something other than a pattern of cumulative error. That is, additional final errors in sentence-nonfinal IUs may represent interference from elements that occur on the other side of the sentence-internal boundary. This possibility seems especially likely when one looks at the proportion of unit-final errors of anticipation versus perseveration in sentence-final and -nonfinal IUs in Figure 3. In particular, nearly half of the unit-final errors in sentence-nonfinal IUs were errors of anticipation ($N = 95$) and a smaller proportion were errors of anticipation and perseveration ($N = 18$), whereas all of the unit-final errors in sentence-final IUs were errors of perseveration ($N = 90$).

By definition, anticipatory and perseveratory errors take their source from a subsequent or prior element in the sentence. In the present data set, only 6% of these sequential errors took their source from within the word. The majority took their source from either an adjacent word or a non-adjacent word within the sentence (57% and 37%, respectively). This means that unit-final errors of anticipation in sentence-nonfinal IUs almost always took their source from a word on the other side of the sentence-internal boundary.

The same distribution of unit-final errors of anticipation and perseveration occurred in IUs with less than 4 words. Half of unit-final errors in smaller sentence-nonfinal IUs were due to interference from elements in the subsequent IUs (unit-final errors in sentence-nonfinal IUs, anticipatory vs. perseveratory errors, $N = 33$ vs. $N = 35$; unit-final errors in sentence-final IUs, anticipatory vs. perseveratory errors, $N = 0$ vs. $N = 40$).

A final set of analyses investigated whether the overall results were qualitatively similar across speakers and sentence types. Since the number of errors per speaker was very small, we compared the number of errors that a speaker produced in the first half of an IU to the number s/he produced in the latter half of the IU. This comparison indicated that 37 out of 40 participants produced more errors in the latter half of an IU compared with the initial half, consistent with the cumulative error pattern and final position effect. Of the 3 participants who did not show this pattern, one had no sequential errors, another had only 2 errors (one in an early IU position and one in a final IU position), and the third had 8 errors (1 in initial IU position, 4 in early IU position, 2 in late IU position, and 1 in final IU position).

The error distribution in non-tongue twister sentences was qualitatively similar to that of tongue twister sentences. Specifically, and in spite of the small number of total errors ($N = 56$), a cumulative error pattern emerged: there were 9 errors in initial position, 9 in early position, 23 in late position, and 15 in final position. The distributional error patterns were also qualitatively similar to those described for tongue twisters when errors in non-tongue twister sentences were investigated in sentence-nonfinal and sentence-final IUs, even though such splitting further reduced the overall number of errors to be investigated. Specifically, there was a final peak in errors for sentence-nonfinal IUs (8 errors in initial position, 4 in early, 6 in late, and 14 in final) and a late peak in errors for sentence-final IUs (1 error in initial position, 5 in early, 17 in late, and 1 in final position).

4. General discussion

The distribution of sequencing errors within multi-word prosodic units was examined in the current study to test for positional effects on errors and to gain insight into how such effects might emerge during speech planning processes. Speech error patterns were found to correspond well with the boundaries of the larger and smaller multi-word prosodic units. In particular, the number of sequencing errors were found to vary as a function of position within a unit such that the fewest errors were found in initial position, more occurred in early-mid position, and even more occurred in late-mid position. We referred to this result as a pattern of cumulative error.

The study also revealed a final position effect: when IUs occurred in sentence-initial or sentence-internal position, a disproportionate number of errors occurred in IU-final position. When IUs occurred in sentence-final position, then the number of errors associated with IU-final position was somewhat reduced relative to the number of errors found in late IU position. The fact that the final position effect varied in this way with sentence position suggests that the cumulative error pattern and final position effect were distinct. In this section, we discuss what each might imply for speech planning.

We propose that the cumulative error pattern provides evidence that the IU represents a planning domain. Croot et al. (2010) found a very similar pattern of results in their study of prosodic prominence and speech errors (see, e.g., their Figure 1). They suggested that the reduced number of errors in utterance-initial position was due to prosodic structure, but could not be sure given their methods. Our results confirm that positional effects on errors relates to prosodic units smaller than the utterance/sentence because the same effects also occur in sentence-medial and -final position. However, the positional effect does not appear to be categorical and associated only with initial position. Rather, it appears to be gradient in that errors are distributed asymmetrically across the prosodic unit. Characterizing the effect as categorical or gradient has consequences for explaining how the errors might occur. Whereas a

categorical effect on errors is well described in a model that scans tagged items, (mis-)selects one, and copies it into its designated slot within a prosodic frame; a gradient effect on errors is better described in a model where the differential activation of linked elements affects selection accuracy.

A gradient pattern of increasing errors across a unit is nonetheless difficult to explain within current network models of speech production. The consensus view is that the serial activation of elements within a plan has more to do with future states than with past ones (Dell et al. 1997). This view is based on the established finding that more errors occur in syllable- and word-initial position than in any other position (e.g., Shattuck-Hufnagel 1983; Dell et al. 1993), and on the documented bias in most populations towards anticipatory speech errors (Dell et al. 1997). For example, Dell and colleagues argue that the bias towards anticipatory errors reflects the more difficult task of planning for the execution of future elements compared to the ease with which speakers are able to inhibit the past.

If future planning really is more difficult than the execution of current elements or the inhibition of past elements, then errors should be more frequent at the beginning of IUs rather than at their end. But this hypothesis is based on the assumption that only the serial activation of elements within a domain is relevant to explaining the distribution of errors across the unit. If we assume that activation of the IU domain itself interacts with the serial activation of elements in its domain, then we might be able to explain the cumulative error pattern without contradicting the view that the future matters more than the past in speech planning and execution.

Specifically, let us assume a temporal window for IU activation and that activation decays during this window. Let us further assume that the relative activation of the IU affects sequencing operations within the IU via cascading activation. As IU activation decays, the degree to which elements on lower levels are differentially activated may also be diminished. Diminished differential activation of serially planned elements could affect sequencing in

speech, resulting in an increase of anticipatory and perseveratory errors. In this way, low error rates on initial words within an IU could reflect a state of high IU activation, and higher error rates on subsequently occurring words the comparably lower state of IU activation.

An alternative possibility is that the cumulative error pattern has little to do with activation levels and instead represents an artifact of using a tongue twister task to elicit speech errors. This possibility seems especially likely when one considers that tongue twisters are difficult because of the frequent alternation between similar sounds in the same structural position across words in a sentence. This alternation creates confusion in access and articulation, a confusion that presumably builds across a sentence as words with similar sounds are repeatedly accessed and articulated. As there were insufficient numbers of anticipatory and perseveratory errors in non-tongue twister sentences to definitively establish whether or not the cumulative error pattern generalizes to these more natural sentences, we cannot discount the possibility that the cumulative error pattern is artifactual. On the other hand, the pattern was noted within IUs in sentences composed of more than one IU; that is, the pattern was repeated in long tongue twisters. So even if the cumulative error pattern is due specifically to the tongue twister's alternation of similar sounding words, this confusion builds across the length of an IU, not across the length of an entire tongue twister. This again suggests that the IU represents a planning domain.

Whereas the cumulative error pattern may reflect decaying activation in the IU that is currently being encoded, the final position effect likely reflects the activation of a subsequent IU-sized domain for phonological encoding. Specifically, we suggest that the process of inhibiting the past, activating the present, and planning the future—to paraphrase Dell et al.'s (1997) formulation of serial language production—occurs at the level of the IU and that the cumulative error pattern and IU-final effects emerge because of this. Final words in sentence-nونfinal IUs are especially error prone because their activation is confounded by the low level

of activation within the IU domain and by the concurrent activation of a subsequent IU-sized sequence of words that is beginning to be phonologically encoded.

Like the cumulative error pattern, it is possible that the final position effect was an artifact of our procedures rather than a phenomenon that lends insight into the planning processes. In particular, the finding that sentences with errors had more prosodic boundaries than sentences without errors might suggest that boundaries follow from errors rather than the other way around. This possibility cannot be discounted in the current study, but seems unlikely because the size of the IUs in sentences without errors was very similar to the size of the IUs in sentences with errors, which suggests coherence in prosodification across sentences. Moreover, the comparison of boundary frequency in sentences with and without errors often depended on the selection of different sentence structures. That is, the prosodification of a medium tongue twister sentence that elicited many errors, like "*Which wristwatches are Swiss wristwatches?*", would frequently have been compared against the prosodification of a medium length tongue twister sentence that elicited fewer errors, like "*The myth of Miss Muffet is famous.*" In other words, harder tongue twisters were often compared against easier tongue twisters. Given this, an alternative interpretation of the result that sentences with errors had more boundaries than those without is that speakers more thoroughly chunk sentences that they perceive will be difficult to execute.

To summarize, the cumulative error pattern and final position effects are interpreted here as evidence for the IU as the principal planning domain for phonological encoding in a hierarchically structured speech plan; and for the idea that the serial production of speech depends on the gradient activation of elements within this domain. We are less clear, though, on whether the IU represents one or two planning domains.

Two levels of boundary strength were coded in the present study and the units defined by these boundaries were referred to as big and small IUs. The identification of larger and smaller units suggests a prosodic hierarchy akin to those described by phonological theories of

intonation where smaller units are embedded within larger ones. Such a suggestion leads to the possibility that the big IU represents the highest level of planning in the hierarchically structured speech plan, and the small IU the next highest level. This possibility is not, however, indicated by the present results. The patterns of speech errors in big and small IUs were deemed so similar that the errors from each were combined in subsequent analyses, preventing any definitive conclusion as to their separate identities. Moreover, the basic similarity between big and small IUs might mean that the distinction between larger and smaller prosodic units is not relevant to planning for the serial production of speech. Big and small IUs could represent the same unit encoded under different conditions. Specifically, it could be that strong boundaries are perceived when encoding of a subsequent IU is delayed until the current IU has been fully encoded, and that weak boundaries are perceived when encoding begins in a subsequent IU before it has ended in the current one.

Alternatively, small IUs may in fact represent real sub-domains of phonological encoding, as suggested by the nested structure of phonological hierarchies. The best way to investigate this possibility would be to look to languages other than English. For example, Korean and Japanese are reported to have accentual phrases (APs) that, unlike the intermediate phrases of English, differ from the larger intonational phrases (IPs) of the language (Beckman and Pierrehumbert 1986; Jun 1998). For example, Korean APs are defined solely by a tonal pattern, whereas Korean IPs (like both the Intermediate and Intonational Phrases of English) are defined by the presence of prosodic boundary cues. If only a single high-level IU is relevant to speech planning, then we might expect to find a relationship between speech errors and AP boundaries or between speech errors and IP boundaries in Korean, but not both. If both units are relevant to speech planning, then we might find differences in the distribution of speech errors within APs and IPs.

Finally, it must be acknowledged that the speech we investigated here may be quite different from speech that is spontaneously generated in everyday conversations. Not only do

the decontextualized stimulus sentences of this and other experimental studies represent a special kind of speech, but the manner in which they were presented may encourage a special kind of speech planning. Here, as in most other experimental investigations of speech production, participants were encouraged to read sentences that were written in normal orthography with punctuation. Participants were also explicitly told to prepare their speech before they spoke. It is possible that tasks such as these encourage planning of longer stretches of speech than that which normally occurs in every day conversations. Such a possibility could be tested by examining error patterns in spoken language corpora.

ACKNOWLEDGEMENTS

We are grateful to Rachel Crist for helping to identify speech errors in these recordings, and to Karen Croot, Mary Huffman, Volya Kapatsinski, Cecilia Kirk, and an anonymous reviewer for commenting on previous versions of the paper. This research was supported in part by Award Number R01HD061458 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD). The content is solely the responsibility of the authors.

APPENDIX A

I. Short Sentences

a. Tongue Twisters

- Please pay promptly.
- Greek grapes are great.
- Fat frogs fly past fast.
- Fred threw three free throws.
- Beth believes thieves seize skis.
- Tighten these knapsack straps.
- Sly Sam slurps Sally's soup.
- One smart fellow, he felt smart.
- Friendly Frank flips fine flapjacks.
- Check the Unique New York shop.

b. Non-Tongue Twisters

- Please pay on time.
- Greek food is great.
- Most frogs can jump high.
- Matt had two free throws.
- People think thieves are crooks.
- The blue knapsack was small.
- The dog slurped the yogurt.
- He's a jolly good fellow.
- Glenwood makes apple flapjacks.
- Check Trader Joes for cookies.

II. Medium Sentences

Speech errors in multi-word prosodic units

a. Tongue Twisters

- Cedar shingles should be shaved and saved.
- This shop stocks socks with stripes and spots.
- Which wristwatches are Swiss wristwatches?
- The myth of Miss Muffet is famous.
- Three grey geese in the green grass grazing.
- She sifted thistles through her thistle-sifter.
- Nine nice night nurses are nursing nicely.
- Six shimmering sharks sharply striking shins.
- Don't pamper damp scamp tramps that camp under ramp lamps.
- Lesser leather never weathered wetter weather better.

b. Non-Tongue Twisters

- We must save our environment.
- The socks he wears have dots and flowers.
- I never wear a fancy wristwatch.
- All the students should read the Greek myths.
- Geese are completely vegetarian.
- John brought thistles to the botany class.
- Nurses are responsible for the sick.
- This magazine has some striking photos.
- Students should take care of spirit lamps in the lab.
- Weather's played an important role in human history.

III. Long (15 to 25 syllables)

a. Tongue Twisters

- The crow flew over the river with a lump of raw liver.

Speech errors in multi-word prosodic units

- I correctly recollect Rebecca MacGregor's reckoning.
- A big black bug bit a big black bear, making the big black bear bleed blood.
- I slit the sheet, the sheet I slit, and on the slitted sheet I sit.
- Give papa a cup of proper coffee in a copper coffee cup.
- A skunk sat on a stump and thunk the stump stunk, but the stump thunk the skunk stunk.
- He would chuck, he would, as much as he could, and chuck as much wood as a woodchuck should.
- I saw Susie sitting in a shoe shine shop: where she sits she shines and where she shines she sits.
- If Peter Piper picked a peck of pickled peppers, where's the peck of pickled peppers Peter Piper pickled?
- While we were walking, we were watching window washers wash Washington's windows with warm washing water.

b. Non-Tongue Twisters

- Cod liver oil is a great source for vitamin A and D.
- He recollected the story of Farmer John and the llama.
- Susie happily watched polar bears at the zoo sliding on ice.
- Slitting the cloth into strips, she sat prettily in the corner.
- Copper mines are depleted, so copper wiring's become expensive.
- She whispered a tune while sitting on the toilet in a noisy bathroom.
- I had a party last night, so I spent the day washing dishes and chucking trash.
- The sun was high in the sky and shining beautifully when the little boy went out walking.
- The people in this town came together to pickle vegetables in preparation for winter.
- The basic and important practice of washing hands before each meal is no longer enforced in schools.

REFERENCES

- Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in English and Japanese. *Phonology Yearbook* 3, 255-309.
- Bolinger, Dwight. 1989. *Intonation and its uses*. Stanford, CA: Stanford University Press.
- Byrd, Dani & Elliot Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31, 149–180.
- Cho, Taehong & James M. McQueen. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33, 121-157.
- Croot, Karen, Claudia Au & Amy Harper. 2010. Prosodic structure and tongue twister errors. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio, & Nathalie Vallee (eds.), *Laboratory Phonology 10*, 433-461. Berlin & New York: Mouton de Gruyter.
- Crystal, David. 1969. *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
- Cutler, Anne, Delphine Dahan & Wilma van Donselaar. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40, 141–201.
- Dell, Gary S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93, 283-321.
- Dell, Gary S., Cornell Juliano & Anita Govindjee. 1993. Structure and content in language production: a theory of frame constraints in phonological speech errors. *Cognitive Science* 17, 149-195.
- Dell, Gary S., Lisa K. Burger & William R. Svec. 1997. Language production and serial order: a functional analysis and a model. *Psychological Review* 104, 123-147.
- Ferreira, Fernanda. 1993. Creation of prosody during sentence production. *Psychological Review* 100, 233-253.
- Fougeron, Cécile & Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic

- domains. *Journal of the Acoustic Society of America* 101, 3728-3740.
- Frazier, Lyn, Katy Carlson & Charles Clifton Jr. 2006. Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences* 10, 244-249.
- Fromkin, Victoria A. 1971. The nonanomalous nature of anomalous utterances. *Language* 47, 27-52.
- Goldrick, Matthew & Sheila E. Blumstein. 2006. Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes* 21, 649-683.
- Goldrick, Matthew. 2007. Connectionist principles in theories of speech production. In M. Gareth Gaskell (ed.), *Oxford Handbook of Psycholinguistics*, 515-530. Oxford, England: Oxford University Press.
- Gussenhoven, Carlos. 2002. Phonology of intonation. *Glot International* 6, 271-284.
- Jun, Sun-Ah. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15, 189-226.
- Keating, Patricia A. & Stefani Shattuck-Hufnagel. 2002. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics* 101, 112-156.
- Krivokapic, Jelena. 2007. Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics* 35, 162–179.
- Ladd, D. Robert. 2000. Tones and turning points: Bruce, Pierrehumbert, and the elements of intonational phonology. In Merle Horne (ed.), *Prosody: Theory and experiment—studies presented to Gösta Bruce*, 37–50. Dordrecht: Kluwer.
- Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, Willem J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1-75.
- MacKay, Donald. G. 1987. *The Organization of Perception and Action: A Theory for Language and Other Cognitive Skills*. New York: Springer.

Speech errors in multi-word prosodic units

- Nespor, Marina & Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.
- Selkirk, Elisabeth. 1995. Sentence Prosody: Intonation, Stress, and Phrasing. In John. A. Goldsmith (ed.), *The Handbook of Phonological Theory*, 550-569. Cambridge: Blackwell.
- Shattuck-Hufnagel, Stefani. 1979. Speech errors as evidence for a serial-ordering mechanism in sentence production. In William E. Cooper & Edward C. T. Walker (eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, 295-342. Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, Stefani. 1983. Sublexical units and suprasegmental structure in speech production planning. In Peter F. MacNeilage (ed.), *The production of speech*, 109-136. New York: Springer.
- Shattuck-Hufnagel, Stefani & Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25, 193-246.
- Stemberger, Joseph P. 1985. An interactive activation model of language production. In Andrew W. Ellis (ed.), *Progress in the Psychology of Language*, 143-186. Erlbaum, Hillsdale, NJ.
- Tabain, Marija. 2003. Effects of prosodic boundary on /aC/ sequences: Articulatory results. *Journal of the Acoustical Society of America* 113, 2834-2849.