CAMBRIDGE
UNIVERSITY PRESS

Cumulative exposure to fast speech conditions duration of content words in English

Earl Kjar Brown

Brigham Young University, USA
Email: ekbrown@byu.edu

Abstract

This paper tests the idea that the speech rate with which surrounding words are spoken affects the mental representation of words and conditions production of words. This possibility is operationalized by measuring a word's ratio of occurrence in speaker-relative fast speech. Other variables shown in the literature to influence speech rate are controlled for in a 10,000-iteration bootstrapping procedure of a mixed-effect linear regression model. The results of the analysis of 39,397 tokens of content words from 1,232 word types in English display a significant effect for a word's ratio of conditioning in speaker-relative fast speech, although the effect size is small or very small. Other variables shown in the literature to condition speech rate also significantly condition speech rate here. This paper suggests that in addition to other aspects of the context of use of words, contextual speech rate also influences the mental representation of words.

Speech rate fluctuates as a function of multiple factors. One conditioning variable is the predictability of words given surrounding words (see Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999; Jurafsky, Bell, Gregory, & Raymond, 2001). Using the Switchboard Corpus (Godfrey, Holliman, & McDaniel, 1992), Bell, Brenier, Gregory, Girand, and Jurafsky (2009) showed that words are pronounced more quickly when they are predictable given the following word. They also found that content words (i.e., nouns, adjectives, verbs, adverbs) and function words (e.g., determiners, pronouns, quantifiers, prepositions) react differently to frequency, and that an effect from the predictability of words given preceding words is limited to only very frequent function words. Further, function words are spoken more quickly on average than content words, and, among content words, more frequent words are articulated more quickly than less frequent ones. Also, second or later occurrences of content words in a conversation are pronounced more quickly than first mentions. Pluymaekers, Ernestus, and Baayen (2005a) also showed an effect from predictability and repetition in an analysis of the duration of the seven most frequent adjectives with the Dutch suffix *-lijk* ('-ly', '-al', '-ous', and '-able' in English).

In addition to the predictability of words given neighboring words, the position of words in utterances, and the frequency of words modulate speech rate. In their

analysis of 130 years of New Zealand English, Sóskuthy and Hay (2017) found that words' durations lengthened as their rate of occurrence in utterance-final position increased. This result is expected given the evidence that utterance-final position fosters elongation of sounds (Cohen Priva, Edelist, & Gleason, 2017). Conversely, Sóskuthy and Hay showed that word durations decreased in words that became more frequent, another expected result given the evidence of a shortening effect in frequent words (see Aylett & Turk, 2006; Pluymaekers, Ernestus, & Baayen 2005b).

In addition to the previously mentioned factors, speech rate varies as a function of the people with whom speakers interact. Cohen Priva et al. (2017) used the Switchboard corpus to test the idea that speakers adjust their speech rate to converge toward the speech rate of their interlocutors. The authors found that to be the case. Also, speakers changed their speech rate in response to characteristics of their interlocutors. For example, speakers talked more slowly with older people, and males spoke more quickly with other males (see also Pépiot, 2014; Van Borsel & De Maesschalck, 2008). In addition to observational data in the Switchboard corpus, experimental data also provide evidence of a convergence in speech rate between speakers, as Freud, Ezrati-Vinacour, and Amir (2018) found this effect in an experiment with ten adult speakers. They concluded that speech rate convergence was nonlinear and was affected by both linguistic and situational factors (see also Borrie & Liss, 2014; Wynn, Barrett, & Borrie, 2022; Wynn, Borrie, & Sellers, 2018). In addition to an effect from the interlocutor, speech rate of words is also influenced by the length of words. In general, longer words are spoken with a quicker speech rate than shorter words (Bell et al., 2009; Lehiste, 1970).

In summary, multiple influences modulate speech rate in spontaneous and semi-spontaneous speech. These factors include the predictability of words based on neighboring words, the position within the utterance where words occur, the token frequency of words, length of words, and social characteristics of speakers' interlocutors. The purpose of the current paper is to analyze the conditioning effect, if any, of a word's cumulative exposure to speaker-relative fast speech, something that has received little attention in the literature. It puts to test the idea that the global effect of cumulative exposure to fast speech influences word duration even after controlling for the local effect of the contextual speech rate with which words are uttered.

Cumulative exposure to conditioning contexts

A major tenet of usage-based models of language is that the contexts in which words are used affect the mental representation of those words. One level of detail that is posited to be stored with words is the proportion with which words occur in contexts that favor their phonetic modification. A growing body of literature provides empirical support for the idea expressed by Bybee (2002:261): "Words that occur more often in the context for change change more rapidly than those that occur less often in that context." Bybee found support for this notion in an analysis of postconsonant word-final *t/d* deletion in Chicano English in Los Angeles as well as in the Switchboard corpus.

Others have found empirical evidence in support of the idea that cumulative exposure to the conditioning context for a phonetic change conditions higher rates of a

phonetic modification. Eddington and Channer (2010) studied the articulation of prevocalic word-final /t/ in American English (e.g., *it is*), as seen in the Santa Barbara corpus (Du Bois, Chafe, Meyer, Thompson, & Martey, 2003). The authors found higher rates of glottalization (i.e., /t/ > [ʔ]) among younger speakers and showed that, word-finally, /t/ was more often followed by consonants, the context favorable to glottalization. Kaźmierski (2020) found further evidence of the importance of cumulative contextual exposure on word-final /t/ glottalization in a corpus of Midland American English: words that frequently occurred in the context for glottalization (i.e., preconsonantally) displayed higher rates of glottalization even when the phonological context was statistically controlled for in a mixed-effect logistic regression.

Seyfarth (2014) studied the duration of words with an eye to the effect of predictability of words given surrounding words. One finding pertinent to our discussion is that words that usually occurred in predictable contexts given surrounding words had shorter duration, even in unpredictable contexts. The author attributed this finding to a lexical effect in storage, that is, words that occurred often in the context for shorter duration were more likely to be pronounced with shorter durations because the cumulative exposure to the conditioning context was stored in memory with the words. Another study with a different variable in English was reported by Forrest (2017). In a large-scale study of the alternation of alveolar [n] and velar [ŋ] pronunciations in word-final -*ing* (e.g., *walking*) among 132 speakers in North Carolina, USA, the author found that frequent occurrence in the phonological contexts that favored the alveolar pronunciation amplified the effect of token frequency. Conversely, the conditioning effect of token frequency was dampened on words that frequently occurred in contexts that favored the velar pronunciation. These results provide evidence of an important interaction between cumulative occurrence in the phonetic contexts for change and the frequency with which words occur.

Turning to Spanish, E. L. Brown (2004) studied the aspiration of word- and syllable-initial /s/ (e.g., *la señora* > *la* [h]*eñora* 'the woman') in colonial US New Mexican and southern Coloradoan Spanish, and found that words that occurred proportionally more often in the context for aspiration, that is, following a nonhigh vowel, had higher rates of aspiration, even outside of that context. E. K. Brown (2009) found further evidence in support of a word's ratio of conditioning in a study of the aspiration and deletion of syllable- and word-final /s/ in Cali, Colombia. E. L. Brown and Raymond (2012) analyzed the variable modern-day outcome in Spanish of word-initial Latin /f/ before a vowel (e.g., Lat. *favor* > Span. *favor* 'favor'; Lat. *facere* > Span. *hacer* 'to do, to make')[1] by analyzing the Medieval Spanish play *La Celestina* published in 1499. Those authors discovered a significant correlation between rates of occurrence of initial-/f/ words after nonhigh vowels and rates of deletion in modern-day Spanish, and they argued for the preeminence of word's ratio of conditioning over token frequency. In another study of /f/ in Spanish, E. K. Brown and Alba (2017) found similar results in an analysis of word-initial /f/ in immigrant Mexican Spanish in California. Additionally, E. K. Brown (2020) studied the variable voicing of word-final /s/ in Spanish (e.g., *lo*[z] *niños* 'the children, the boys'), and found that words that occurred relatively often in the context for voicing (i.e., before a voiced consonant) showed higher rates of voicing than other words.

While evidence is mounting about the influence of cumulative exposure to phonological contexts that condition a phonetic change, little has been said in the literature about the effect of cumulative exposure to speech rate. Does cumulative exposure to the speech rate of utterances have an effect on word durations? Are words that occur often in fast speech articulated more quickly than other words, even when token-specific contextual speech rate is controlled for? Only E. L. Brown, Raymond, Brown, and File-Muriel (2021) investigated the effect of cumulative exposure to fast speech. Those authors studied the speech rate of words and the duration of the /s/ segment in a sample of the Spanish of Cali, Colombia. Those authors measured the contextual speech rate of each word by creating a continuous variable of phones per second based on the stretch of speech between the target word and the end of the utterance. They split their dataset approximately in half and used one subset of the data to calculate the average rate of speech of the speakers in the study, as well as the proportion with which words occurred in speech that was faster than average for the speakers who articulated the words. With the other half of the dataset, they modeled the effect of this proportion (i.e., a word's ratio of conditioning), along with other predictor variables shown in the literature to condition speech rate. Their findings showed that words that occurred more often in speech that was faster than average for the speakers who spoke them displayed a quicker speech rate in comparison to other words, even when the token-specific speech rate was statistically controlled for. This finding adds credence to the assertion that the contexts of use, including the contextual speech rate with which words are spoken, have an effect on the mental representation of words.

Research questions

This paper seeks to help fill the gap in the literature about the effect of cumulative exposure to speech rate by testing E. L. Brown et al.'s (2021) findings in a different language. This paper utilizes the Buckeye Corpus (Pitt, Dilley, Johnson, Kiesling, Raymon, Hume, & Fosler-Lussier, 2007) of English in central Ohio, USA, and operationalizes the cumulative exposure to speech rate by measuring a word's ratio of conditioning in speaker-relative fast speech. The research question is:

> RQ: What effect, if any, does a word's ratio of conditioning in speaker-relative fast speech have on word duration in a sample of American English?

The hypothesis is that the speech rate in which words are used in conversational language influences the mental representation of words, and consequently, future production of those words.

Materials and methods
Corpus

In order to test the research question, the Buckeye Corpus (Pitt et al., 2007) was accessed. This corpus contains a collection of interviews with forty speakers who conversed freely with an interviewer in a modified sociolinguistic interview (Labov, 1984;

Tagliamonte, 2006). The interviewers engaged the participants in conversations about everyday topics, such as politics, sports, traffic, and schools. Most speakers belonged to the middle socioeconomic class or the upper working class. The process of making the recordings was completed between fall 1999 and spring 2000 and the corpus comprises about 300,000 words. The project to create the corpus received approval from the Internal Review Board of the Ohio State University (see Pitt et al., 2007, for details).

Variables

The response variable under investigation in this study is the duration of words measured in seconds. The duration of each word token was calculated as the sum of the durations of the articulated sound segments in the word, as noted in the phonetic transcription provided in Buckeye. It should be noted that the creators of Buckeye considered long vowels with glides (e.g., the vowels in *boy*, *boat*, *bait*) as single segments.

The predictor variable of interest in this study is cumulative exposure to speech rate, here operationalized as a word's ratio of conditioning (WRC) in speaker-relative fast speech.[2] This measure was calculated with the following steps. First, the contextual speech rate of each token was measured by taking the number of segments spoken between the target word and the end of the utterance (excluding the target word itself), and dividing that number by the duration in seconds of that stretch of speech. For example, during their interview, Speaker 8 said this phrase: *Actually, I think it needs to probably start younger than that*. When considering the verb *needs* as the target word, the contextual speech rate was calculated by taking the number of segments articulated in *to probably start younger than that* (twenty-three segments in the phonetic transcript) and dividing it by the sum of the durations of those segments (1.508 seconds), resulting in a contextual speech rate of this particular token of *needs* of 15.25 segments per second (i.e., $23/1.508 \simeq 15.25$). Next, the average speech rate of each speaker was calculated by taking the number of segments articulated by the speaker in the interview and dividing it by the number of seconds that the speaker took to pronounce those segments. For example, Speaker 8 articulated 18,804 segments, and the sum of the durations of those segments is 1,432.68 seconds, which gives that speaker an average speech rate of 13.13 segments per second (i.e., $18,804/1,432.68 \simeq 13.13$). Third, the relative speed of the contextual speech rate of each target word was compared to the average speech rate of the speaker who produced it. Target words whose contextual speech rate was articulated faster than the average speech rate of the speaker were labeled "fast," while other words were labeled "slow." To continue with the above example, that particular token of *needs* spoken by Speaker 8 was labeled "fast" because its contextual speech rate (i.e., 15.25 seg./sec.) is faster than the average speech rate of that speaker (i.e., 13.13 seg./sec.). Finally, the proportion of tokens of each word type that occurred in speaker-relative fast speech was calculated, and that proportion was the WRC measure for the word type. To take another example, the word *honest* occurs fourteen times across all forty interviews, with six of these tokens occurring in speaker-relative fast speech and eight tokens in slow speech. Thus, the WRC measure for *honest* is the quotient of six divided by fourteen, that is, 0.429.

Other predictor variables shown in the literature to condition speech rate were also accounted for. Of course, word length influences word duration, as, all things being equal, words with more segments take longer to pronounce than words with fewer segments. However, it has also been shown that, on average, segments in long words are spoken more quickly than segments in short words (see Bell et al., 2009; Cohen Priva, 2017). For example, Lehiste (1970) shows that the three segments in *luck* become successively shorter in *luck, lucky, luckily*. It should be noted that the phonetic realization of each token, not the phonemic structure, was utilized to measure word length. To illustrate, despite the fact that the word *honest* phonemically has five segments (i.e., [anɪst] or [anʌst]), in the dataset it was sometimes pronounced with five segments and sometimes with four segments (i.e., [anɪs], [anʌs], [ãɪst], [aŋɪz]). Consequently, some tokens of *honest* have a word length of five segments while others have a word length of four segments.

Word durations are influenced by token-specific contextual speech rate, that is, words spoken in fast utterances should themselves be spoken quickly and therefore have shorter durations. As such, the contextual speech rate of the words following each target word within the same utterance was calculated as a continuous variable of segments per second. In her study of homophones in English, Gahl (2008) showed that contextual speech rate conditioned word durations. The author operationalized this variable by analyzing the speech rate of the stretch of speech before each target word, as well as the stretch of speech after the target word, within the same pause-bounded utterance. However, Gahl did not find that preceding speech rate (i.e., the speech rate of the words before the target word) made a significant contribution to the prediction of target word duration in her data. In contrast, she found a significant conditioning effect from the speech rate of words following the target word within the same utterance. Likewise, E. L. Brown et al. (2021) showed a significant effect from posttarget speech rate on word-level speech rate as well as on the duration of /s/ in a sample of Colombian Spanish. In the current paper, contextual speech rate was also operationalized by measuring the speech rate of words following target words within the same pause-bounded utterance. This variable describes the local context-level speech rate that a specific token happens to be used in, while WRC accounts for the global or cumulative exposure of words to context-level speech rate. It should be noted that while exploring the data for the current paper, contextual speech rate before target words was also measured, but like in Gahl's and E. L. Brown et al.'s studies, no significant effect was found from this pretarget speech rate, and hence only posttarget speech rate is used here.

Other predictor variables included the distance to the end of the utterance and the length of utterances. Sounds are elongated at the end of utterances, and, as such, words closer to the end of utterances are more likely to have longer durations than words farther from the end.[3] Similarly, the length of utterances was measured as a continuous variable of the number of words. Studies have found that speech rate increases as utterance length increases, that is, words in longer utterances are pronounced more quickly than words in shorter utterances, all things being equal (see Cohen Priva et al., 2017; Jacewicz, Fox, & Wei, 2010).

Several predictor variables dealing with the resting state of mental activation of words as well as frequency were coded for. In order to control for a possible priming

of words, whether the target word had been previously mentioned by the speaker was marked. Concerning token frequency, on average more common words are pronounced more quickly than rare words, and consequently, token frequency was accounted for. The token frequency of words was measured in the OpenSubtitles English corpus (Lison & Tiedemann, 2016) in the ten-year period ending in 2000, that is, subtitles of movies and TV shows released only between the years 1991 and 2000 were used. This time period was deliberately chosen in order to use word frequencies likely representative of the ten-year period ending with the collection of the Buckeye Corpus in fall 1999 and spring 2000. This subcorpus contains more than 292 million words. In the statistical analysis reported below, frequencies were transformed with Laplace transformation, as some of the target words were not attested in the OpenSubtitles subcorpus. This procedure adds one to each frequency count, increases the corpus size by the number of word types in the corpus, and then calculates the relative frequency of each word with these adjusted figures (see Brysbaert & Diependaele, 2013, for details).

Another measure of cumulative experience in language is predictability of target words given surrounding words. In this paper, the directional predictability measures forward Delta P (ΔP forward) and backward Delta P (ΔP backward) are employed (Schneider, 2020). These scores are based on conditional probability (see Jurafsky et al., 2001), but include a "small adjustment, which 'punishes' pairs whose second word also frequently occurs in other combinations" (Schneider, 2020:255).[4] To illustrate these predictability measures, the forward Delta P of *idea* given *faintest* (i.e., *faintest idea*) is high, at 0.72, while the forward Delta P of *kind* given *still* (i.e., *still kind*) is low, at 0.00001. In other words, speakers are relatively likely to say *idea* after *faintest*, but do not tend to say *kind* after *still*. As additional examples, the backward Delta P of *supreme* given *court* (i.e., *supreme court*) is high, at 0.42, while the same measure of *last* given *experiments* (i.e., *last experiments*) is low, at 0.00003. This is to say that speakers are relatively likely to say *supreme* before saying *court*, but do not tend to say *last* before *experiments*. It should be noted that the Delta P scores in these data are in units of bits of information, based on Laplace transformed frequencies, by taking the negative logarithm to base 2 of frequencies in order to deal with the logarithmic nature of frequencies (Cohen Priva, 2017). Also, the frequency of the first "word" in the calculation of backward Delta P for utterance-initial target words was taken to be the transformed number of utterances in the subcorpus, which serves as a proxy for the number of pauses.

Two extralinguistic predictor variables were accounted for: age and sex of the speakers. It should be noted that the corpus designers grouped speakers based on age: "young" speakers were younger than thirty years old, while "old" speakers were older than forty. As such, rather than a continuous variable based on age in years, age was entered as a binary categorical variable in the statistical analysis reported below.

Finally, random effects were entered for *speaker* and *word*. Specifically, *speaker* was entered as a random slope with the predictor variable of interest here, a word's ratio of conditioning, and *word* was entered as a random intercept. During the exploratory phase of this paper, *word* was also entered in the statistical model as a random slope with a word's ratio of conditioning, but due to singularity issues, it was removed, and

only a random intercept was kept. The purpose for these two random effects was to control for variability caused by natural difference in the speech rate among the forty speakers, as it is safe to assume that some speakers speak more slowly than others, while others speak faster. Likewise, differences in word durations may be attributable to individual words, that is, some words may simply be spoken more quickly than other words, regardless of the conditioning effect of other variables.

Data exclusion

A series of exclusions reduced the number of tokens entered into the statistical analysis. The durations of only content words (i.e., nouns, verbs, adjectives, and adverbs) is analyzed here, as it has been reported that content words and function words react differently to frequency effects and have different routes of access (see Bell et al., 2009; Segalowitz & Lane, 2000; Seyfarth, 2014). Also, only words whose WRC value was based on ten or more tokens are included. This decision was made in an effort to ensure that WRC scores are the result of systematic patterns rather than arbitrary variability. Put another way, words that occur only a few times may happen to occur in fast speech or they may happen to occur in slow speech, and in order to try to control for this haphazardness, only words whose WRC score was based on ten or more tokens were retained. E. K. Brown (2020) highlights the importance of basing WRC on more than a few tokens in his analysis of word-final /s/ voicing in Mexican Spanish. Words in utterance-final position are also excluded, as of necessity, each token must have following sounds with which to calculate the speech rate based on the stretch of speech from immediately after each target word to the end of the utterance. Filled pauses were then excluded, identified by their orthography (i.e., *um*, *uh*, *um-hum*, *uh-huh*, *aha*, *ah*, *uh-hum*, *um-huh*, *uh-uh*, *uh-oh*, *uh-hmm*) or by the Penn Treebank part-of-speech tag "UH" (which included some tokens of *yes*, *yeah*, *okay*, and *wow*). Additionally, words followed by filled pauses were also excluded from the dataset. In an effort to gain access to speakers' vernacular or a speech variety close to that vernacular, words spoken during the first five minutes of each interview were excluded from analysis, as speakers in general become more relaxed with their speech as an interview progresses. To reduce the skewing effect of outliers, words with a duration that lay outside three standard deviations above or below the mean word duration were excluded. Likewise, words in utterances of three or fewer words were excluded, as these short utterances are often backchannels. This series of exclusions left 39,397 word tokens from 1,232 word types for the statistical analysis.

Statistical analysis

Mixed-effects linear regression was used to measure the influence of the predictor variables on the response variable. Concerning the assumptions of regression, Variance Inflation Factors (VIFs) showed no multicollinearity of the predictor variables, and thus that assumption was met. However, upon visual inspection of the residuals of the model, a nonnormal distribution was evident, thus violating the assumption of normally distributed residuals. Consequently, a Box-Cox power transformation ($\lambda = 0.24$) of the response variable (i.e., word durations) was performed

(see Box & Cox, 1964; Levshina, 2015:158). Also, the assumption of homoscedasticity of residuals (i.e., that the residuals vary constantly) was violated as well. Because of these violations of the assumptions of regression analysis, a 10,000-iteration bootstrapping procedure was performed. Levshina (2015:167) points out that: "When one or more assumptions of linear regression have been violated, one can use regression based on bootstrapping…in order to return results that can be trusted." Details about the violations of the assumptions that led to the decision to bootstrap the linear regression as well as details about the software used are available in the online appendix.

To summarize in list format, the variables entered in the statistical analysis are presented in Table 1.

The data files and *Python* and *R* scripts used in this paper are available at the following Open Science Framework repository:

https://osf.io/dgfj5/?view_only=0c2e3b6458614715a8ad333c3bca6ebb

Results

The results of the statistical analysis reveal significant effects from a handful of predictor variables, including the variable of interest here: a word's ratio of conditioning (WRC). As mentioned above, violations of some of the assumptions of regression analysis motivated a Box-Cox transformation of the response variable and a bootstrapping procedure of the linear regression model. The bootstrapping procedure was performed to create 95% confidence intervals for the coefficients of the model.

Table 2 presents the results of the bootstrapping procedure. For each variable in the rows, four values are given in columns: two dealing with the variable's slope and two dealing with the associated *p*-value. In each set of two values, the lower and upper limits of the 95% confidence interval (i.e., the 2.5[th] and 97.5[th] percentiles) are given. Turning our attention first to the *p*-values, as seen in the table, there are five predictor variables whose upper limit of their 95% confidence interval (i.e., their 97.5[th] percentile) falls below an alpha level of 0.05. They are: word length, speech rate, word frequency, WRC, and forward Delta P. A sixth predictor, previous mention, has a 97.5[th] percentile *p*-value just above 0.05, at 0.0548; the *p*-value of 0.05 falls between the 97.1[st] and 97.2[nd] percentiles. The remaining predictor variables in the model have 97.5[th] percentile *p*-values that fall above 0.05. They are: the interaction term between word frequency and WRC, the distance of a target word to the end of the utterance, the length of the utterance, and backward Delta P (see Table 2). Note that the values are rounded to four digits, which causes some values to be rounded to zero.

We now turn our attention to the slopes of the six predictors whose 97.5[th] percentile *p*-values fall below or slightly above (i.e., previous mention) the alpha level of 0.05. Word length has a positive slope, with 95% confidence intervals (CI) ranging from 0.5565 to 0.5829, indicating that as word length increases, word duration also increases. As expected, the more segments in a word, the longer it takes speakers to articulate that word. This effect is illustrated in Figure 1. Note that the points are jittered and slightly transparent to ease overplotting.

Table 1. Variables entered in the statistical analysis

Variable	Type	Description
Word duration	Response	Box-Cox transformed values based on number of seconds (z-scores)
Speaker	Random slope with WRC	40 speakers
Word	Random intercept	1,232 word types
Age	Fixed	Young (<30 years old), old (>40 years old) (z-scores)
Sex	Fixed	Female, male
Word's ratio of conditioning (WRC)	Fixed	Proportion of occurrences of word type in speaker-relative fast speech (z-scores)
Token frequency	Fixed	Log of Laplace-transformed values based on counts from OpenSubtitles corpus (1991–2000) (z-scores)
Interaction between WRC and token frequency	Fixed	Included given Forrest's (2017) findings (z-scores)
Post-target contextual speech rate	Fixed	Phones per second (z-scores)
Distance to end of utterance	Fixed	Number of words (z-scores)
Previous mention	Fixed	Whether the target word was previously mentioned in the interview: Yes, No
Utterance length	Fixed	Number of words (z-scores)
Forward predictability	Fixed	Forward Delta P of target word given previous word (z-scores)
Backward predictability	Fixed	Backward Delta P of target word given following word (z-scores)
Word length	Fixed	Number of segments (z-scores)

The slopes from the bootstrapping procedure for speech rate are negative (95% CI ranging from −0.1194 to −0.1056), indicating a negative relationship: as speech rate increases, word duration decreases, as expected (see Figure 2).

The effect of word frequency also appears to significantly condition word duration, as illustrated in Figure 3. With negative slope coefficients (with lower and upper 95% CI intervals of −0.0666 and −0.0379), the higher the frequency of words, the shorter their durations, as expected.

The predictability of the target word given the preceding word (i.e., forward Delta P) has a positive relationship with word duration, as the lower and upper 95% CI of the bootstrapped slope coefficients are positive (i.e., 0.0256 and 0.0396). Note that the scale of forward Delta P is transformed into bits of information, and the results are interpreted as follows: the more information that a word contributes

Table 2. The lower and upper limits of the 95% confidence intervals (i.e., 2.5th and 97.5th percentiles) of the coefficient estimates of slopes and *p*-values from 10,000 bootstraps

Variable	Slope		*p*-value	
	2.5th percentile	97.5th percentile	2.5th percentile	97.5th percentile
(Intercept)	0.2184	0.2507	0.0000	0.0000
Word length	0.5565	0.5829	0.0000	0.0000
Speech rate	−0.1194	−0.1056	0.0000	0.0000
Word frequency	−0.0666	−0.0379	0.0000	0.0124
Word's ratio of conditioning (WRC)	−0.0671	−0.0393	0.0002	0.0249
Word frequency*WRC	−0.0323	−0.0147	0.0023	0.1530
Forward Delta P	0.0256	0.0396	0.0000	0.0000
Previous mention: yes	−0.0506	−0.0178	0.0000	0.0548
Distance to end	−0.0115	0.0058	0.0086	0.9678
Utterance length	−0.0140	0.0034	0.0017	0.9505
Backward Delta P	0.0036	0.0195	0.0000	0.3709

(i.e., the inverse of predictability), the longer it takes a speaker to articulate the word, as expected (see Figure 4).

As mentioned above, the alpha level of 0.05 falls just within the upper limit of the 95% confidence interval of the predictor variable that tracked whether the target word was previously mentioned in the interview. Specifically, the value of 0.05 falls between
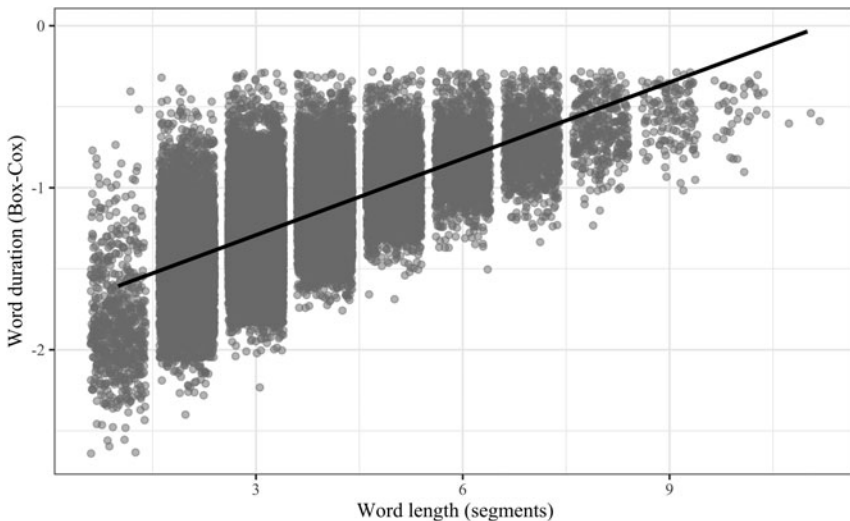
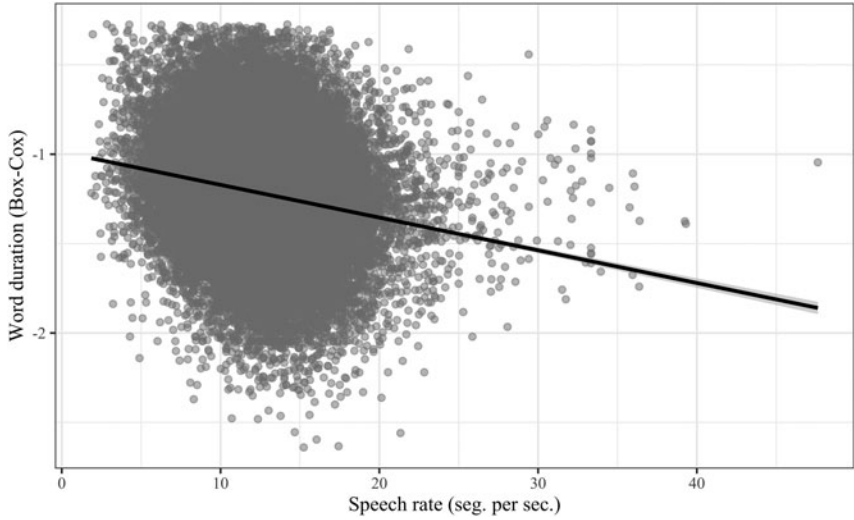**Figure 1.** Word duration by word length.

Figure 2. Word duration by speech rate.

the 97.1st and 97.2nd percentiles of the 10,000 bootstrapped *p*-values for this variable. As such, it may be safe to assert that this predictor also significantly conditions word duration in these data, if only at a marginally significant level. As expected, words that were previously mentioned in the interview have shorter durations than new words. This is shown in Figure 5, which displays a boxplot overlaid on a violin plot.

Concerning the predictor variable of interest here, a word's ratio of conditioning (WRC), the results also show a significant conditioning effect on word duration.
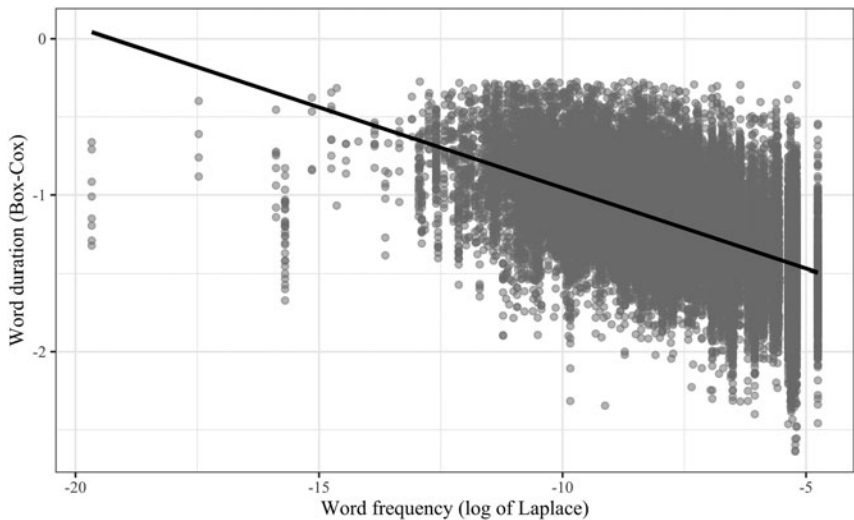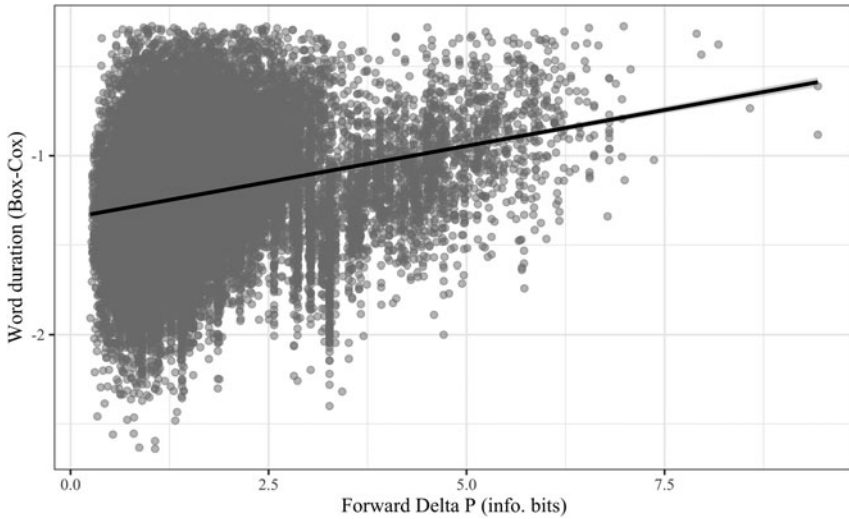

Figure 3. Word duration by word frequency.

Figure 4. Word duration by forward Delta P.

However, the results of the bootstrapping procedure cast doubt on whether the inter-action between WRC and word frequency significantly influences word duration in these data, as the alpha level of 0.05 falls between 72nd and 73rd percentiles of the 10,000 strapped *p*-values of this variable, rather than above the 97.5[th] percentile. In other words, the *p*-value of 0.05 falls well within the 95% confidence intervals, and thus these data do not convincingly demonstrate that the results of the interaction term are not due to random chance.
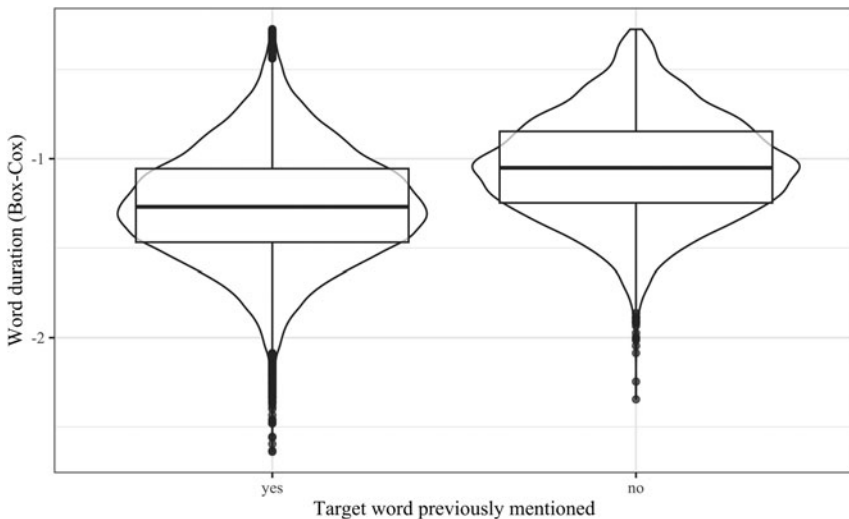

Figure 5. Word duration by previous mention of target word.

In contrast, looking at WRC as a main effect, we see in Table 2 that the 97.5th percentile of p-values (i.e., 0.0249) for this predictor variable falls below 0.05. The negative slope values (95% CI from −0.0671 to −0.0393) indicate a negative relationship, such that as WRC increases, word duration decreases, as expected (see Figure 6).

Turning our attention to the effect sizes of the predictor variables, we see in Table 3 that the predictor variable with the largest effect size, by far, is word length, whose 95% confidence interval falls between 1.5093 and 1.6180, within the range that Sawilowsky (2009) describes as a "very large" effect size. Naturally, the longer words are, the longer it takes speakers to articulate them. Table 3 displays the lower and upper limits (i.e., 2.5th and 97.5th percentiles, respectively) of the 95% confidence intervals of the Cohen D scores for all predictors.

The lower and upper CI absolute values of the Cohen D values for speech rate fall within the range of a "small" effect size, that is, greater than or equal to 0.2 but less than 0.5: lower CI = −0.3577, upper CI = −0.3168. Two predictor variables have lower and upper confidence intervals of Cohen D values that straddle the threshold between "small" and "very small" effects size: word's ratio of conditioning (lower CI = −0.2580, upper CI = −0.1511) and word frequency (lower CI = −0.2455, upper CI = −0.1395). The other predictors with significant effects on word duration given the results of the bootstrapped regression analysis have absolute 95% confidence intervals of Cohen D scores between 0.01 and 0.2 and thus have a "very small" effect size (see Cohen, 1988; Sawilowsky, 2009). To summarize, among the predictor variables with a significant effect on word duration in these data, word length has the largest effect size by far, followed by speech rate, then by WRC and word frequency, and finally by other variables.

Finally, let us inspect the model-level R^2 results. As seen in Table 4, the 95% confidence interval for conditional R^2 ranges from 0.6258 to 0.6426, meaning that
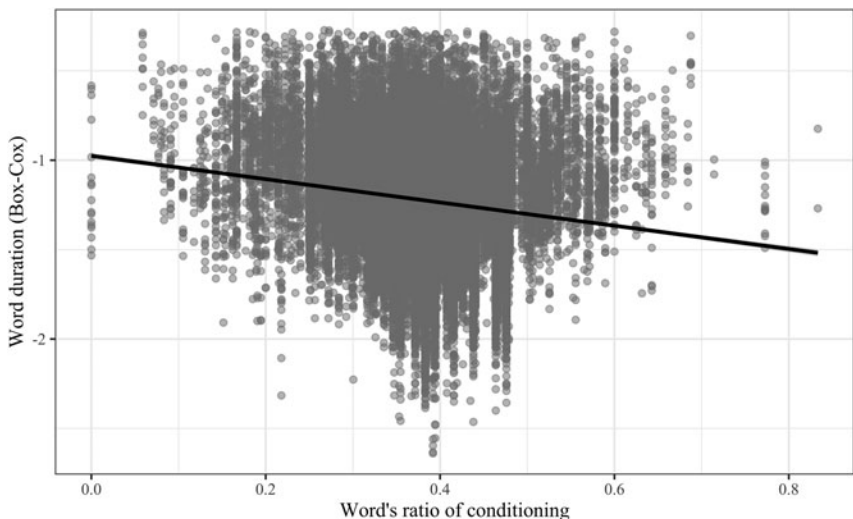

Figure 6. Word duration by a word's ratio of conditioning.

Table 3. Lower and upper limits of 95% confidence intervals of Cohen *D* scores from 10,000 bootstraps

Variable	2.5th percentile	97.5th percentile
Word length	1.5093	1.6180
Speech rate	−0.3577	−0.3168
Word's ratio of conditioning (WRC)	−0.2580	−0.1511
Word frequency	−0.2455	−0.1395
Word frequency*WRC	−0.1755	−0.0832
Forward Delta P	0.0716	0.1109
Previous mention: yes	−0.0553	−0.0194
Backward Delta P	0.0090	0.0494
Utterance length	−0.0318	0.0076
Distance to end	−0.0266	0.0133

Table 4. Lower and upper limits of 95% confidence intervals of conditional R^2 and marginal R^2 values from 10,000 bootstrapped models

Measure	2.5th percentile	97.5th percentile
Conditional R^2	0.6258	0.6426
Marginal R^2	0.4189	0.4375

the combination of the random effects and the fixed effects in the models explain approximately 62-65% of the variability in word durations. The 95% confidence interval for marginal R^2 value ranges from 0.4189 to 0.4375, suggesting that the fixed effects explain approximately 41-44% of the variability in word durations.

In summary, the results of the 10,000-iteration bootstrapping procedure suggest that a handful of predictor variables significantly condition word duration in these data. These predictors include the variable of interest here, that is, WRC. However, the results do not conclusively suggest that the interaction between WRC and word frequency significantly conditions word duration in these data. Nevertheless, there is evidence that WRC as a main effect significantly conditions word duration, despite having a small or very small effect size.

Discussion and conclusions

The results suggest that a word's ratio of conditioning in speaker-relative fast speech conditions word duration in a large sample of content words in the Buckeye Corpus, such that words that are frequently followed by fast speech have shorter durations than other words, even when the speech rate of the moment and other predictors are statistically controlled for. Consequently, the research question can be answered affirmatively. In short, cumulative exposure to contextual speech rate affects word duration. Rather than being an epiphenomenon of other factors, a word's ratio of

conditioning exerts a significant effect, albeit a small or very small one, in these data. This finding suggests that the contexts in which words are frequently used influence later production of words. However, a question remains: How does a word's previous exposure to contextual speech rate affect future production?

One reasonable explanation for this phenomenon is that words are malleable cognitive entities that are sensitive to and affected by the contexts in which they are frequently used. An exemplar representation of linguistic experience (see Pierrehumbert, 2001, 2002, 2003) lends itself as a viable option for explaining how words might be shaped and reshaped by language usage. Bybee (2006:716) proposed that: "A token of linguistic experience that is identical to an existing exemplar is mapped onto that exemplar, strengthening it. Tokens that are similar but not identical … are stored near similar exemplars to constitute clusters or categories." When discussing the deletion of word-final [t] and [d] in American English, Bybee (2010:38) proposed that "the dominance of the cluster by the most frequent exemplar, which thus has a higher likelihood of being chosen for production, leads to the tendency for words to settle on a tight range of variance or a more centered category." All things being equal, an exemplar of a word informed by previous exposure to contextual speech rate is likely to be chosen during production.

Concerning the supposed tightening of the range of variance of a word's exemplar cluster, Wedel (2004, 2006) made an interesting appeal to biological evolution that is useful for our discussion. That author compared linguistic categorization to Darwinian biological evolution and proposed that the tokens of a word exemplar cluster that are most robust are more likely to be selected during production, according to an adaptation of biological natural selection. Additionally, less robust and less frequent tokens of linguistic experience are pruned from the mental representation of the category. Wedel (2006:255) stated that, "If we assume that the activation of exemplar memories decays slowly with time, and that probability of use of an exemplar as a model for production is proportional to its activation, then there will be an effective turnover of exemplars, as old exemplar memories die away and new ones take their place. Whenever an exemplar decays very far before it has a chance to be produced, its line of descent is effectively truncated." Applied to the speech rate of words, we might argue that the most activated exemplars of a word are those that are most frequent in speech and are thus more likely to be naturally selected during production. In contrast, less frequent and less activated exemplars of a word are less likely to be selected and are subject to eventually dying out.

Concerning the exact mechanism by which contextual speech rate influences word duration, it seems that two possibilities exist: (1) contextual speech rate itself is stored with words or (2) contextual speech rate affects the duration of sound segments in words, which are stored in memory. While the data in this study do not explicitly lend themselves to deciphering between these two possibilities, it is posited here that speech rate affects the duration of sound segments, which in turn affect the mental representation of words (i.e., the second option above). Possible evidence in support of this idea is the fact that words can be used outside the contextual speech rate with which they typically occur. In contrast, if the contextual speech rate itself were stored (i.e., the first option above), how would an exemplar of a word from fast speech ever get into slow speech? Are words forever stuck with the contextual speech rate

with which they are used? This explanation feels tenuous at best. Consequently, it may be safe to assert that repeated exposure to fast speech affects the cognitive representation of segments in words, and thus the exemplar clusters of words are molded by contextual speech rate by way of modified sound segments.[5]

Another possible mechanism by which the cognitive representation of words is modified might come from word-specific phonetics. For example, the predictability of segments given other segments within the same word can affect word duration (Seyfarth, 2014). Future research may corroborate this finding in other language data.

Turning our attention to how the current paper relates to other studies, the results presented here concur with those found by E. L. Brown et al. (2021) in Spanish. While their data differ from those used in this paper, their results also suggest that contextual speech rate has an effect on word duration. Both lend empirical support to the idea that cumulative exposure to speech rate affects sound segment durations and word durations. Bybee (2001:52) argued that "mental representations contain considerable detail about phonetic variants, including the specification of multiple acoustic features…." Despite discussing sounds, Bybee's assertion about phonetic variants can likely be applied to words. In addition to phonetic context, morphological context, semantic and pragmatic meaning, and social factors, the results of this study suggest that speech rate influences the cognitive representation of words, likely by way of the modification of sound segments in the words. In more general terms, the results of the current paper—as well as those of E. L. Brown et al. (2021)—suggest that the ways in which language is used affects the structure of language.

Concerning future directions for investigation, some limitations of the current paper include the lack of analysis of a possible conditioning effect from the grammatical constructions that words occur in, and a possible effect from the prefab status of multiword expressions. It might be the case that words are spoken faster (or slower) in certain constructions or when they are in prefabs. Evidence in support of this possibility is seen in a study of vowel duration in English. Bybee and Napoleão de Souza (2019) studied ten adjective types in English in both attributive position (e.g., **hot** weather, **dead** cell phone) and predicative position (e.g., *it's so* **hot**, *my father is* **dead**). They found that vowels in those ten adjective types in attributive position are shorter than they are in predicative position. In addition to grammatical construction, Bybee and Napoleão de Souza (2021) found evidence of the importance of prefab status on the duration of vowels in adjectives in adjective + noun strings (i.e., attributive position). Those authors applied criteria proposed by Erman and Warren (2000) to identify conventionalized multiword strings (i.e., prefabs). Future studies of the cumulative effect of exposure to contextual speech rate on word speech rate would do well to investigate a possible effect from grammatical constructions and prefab status.

Another possible avenue for future research on the effect of a word's ratio of conditioning (WRC) on speech rate is how WRC is operationalized. The current paper follows the methodology of E. L. Brown et al. (2021), which operationalizes WRC based on a binary classification of fast versus slow. Another possibility is to base WRC on a gradient scale of the numeric distance from the average speech rate of the individual speakers. Put another way, WRC could be based on how much faster or how much slower the tokens of a word type are on average in comparison to the average speech rate of the individual speakers.

In conclusion, this paper has analyzed a large sample of content words (39,397 word tokens from 1,232 word types) in the Buckeye Corpus of spontaneous speech from forty speakers in central Ohio, USA. The primary objective was to test the conditioning effect on word duration of cumulative exposure to contextual speech rate, here operationalized as a word's ratio of conditioning (WRC) in speaker-relative fast speech. The results show that a handful of variables significantly condition word duration. Upon investigation of the effect size of significant conditioning variables, it was discovered that the contextual speech rate in the moment had a significant, but small, effect on word duration. As expected, words were articulated more quickly as the token-specific posttarget stretch of speech was pronounced more quickly. Concerning the cumulative exposure to speech rate (i.e., WRC), it seems logical that for this variable to have the possibility of exerting a significant effect on word duration, the contextual speech rate of the moment from which it is derived must first exert a significant influence, and that is the case in these data. As hypothesized, the cumulative exposure to speech rate also significantly conditioned word duration, with shorter word durations correlating with more frequent use of a word in speaker-relative fast speech. While significant, this predictor has a small or very small effect size.

In sum, while this paper utilizes a synchronic corpus of English, the results seem to support the notion expressed by Bybee (2002) that words occurring more frequently in the context for change end up changing more quickly than other words. As suggested above, this effect can be explained if we envision words as malleable cognitive entities that are shaped and reshaped by the contexts in which they are used. In short, this paper proposes that usage affects the mental representation of language, which in turn affects future production.

Notes

1. Orthographic *h* in modern-day Spanish represents phonetic zero in most varieties, outside of the digraph *ch* (e.g., *cancha* 'court, field') and the digraphs *hi* and *hu* when followed by a nonhigh vowel (e.g., *hielo* 'ice', *huerta* 'garden').
2. This variable has various names in the literature, among them the following: Frequency in a Favorable Context (or FFC; E. K. Brown, 2009; E. L. Brown, 2004), discourse context frequency (E. L. Brown, 2015), Frequency in a Reducing Context (E. K. Brown & Alba, 2017), Frequency of use in a Reducing Context (E. L. Brown, 2018), and contextual ratio frequency (E. K. Brown, 2018).
3. It should be noted that Fletcher (2010) reports that phrase-final lengthening is more pronounced at the end of intonational phrases than at the end of utterances. Thanks are expressed to an anonymous reviewer for pointing this out.
4. As expected, Delta P and conditional probability are correlated in these data: forward, $r = 0.95$, $p \leq 0.001$; backward, $r = 0.94$, $p \leq 0.001$.
5. Thanks are expressed to an anonymous reviewer for pointing out this possible mechanism.

References

Aylett, Matthew, & Turk, Alice. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America* 119(5 Pt 1):3048–3058. https://doi.org/10.1121/1.2188331

Bates, Douglas, Mächler, Martin, Bolker, Ben, & Walker, Steve. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1):1–48. https://doi.org/10.18637/jss.v067.i01

Bell, Alan, Brenier, Jason M., Gregory, Michelle, Girand, Cynthia, & Jurafsky, Daniel. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1):92–111.

Borrie, Stephanie A., & Liss, Julie M. (2014). Rhythm as a coordinating device: Entrainment with disordered speech. *Journal of Speech, Language, and Hearing Research* 57(3):815–824. (doi:10.1044/2014_JSLHR-S-13-0149)

Box, George Edward Pelham, & Cox, David Roxbee. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B* 26(2):211–252.

Brown, Earl K. (2009). *A usage-based account of syllable- and word-final /s/ reduction in four dialects of Spanish*. Munich: Lincom Europa.

Brown, Earl K. (2018). The company that word-boundary sounds keep: The effect of contextual ratio frequency on word-final /s/ in a sample of Mexican Spanish. In K. A. Smith & D. Nordquist (eds.), *Functionalist and usage-based approaches to the study of language: In honor of Joan L. Bybee* (Studies in Language Companion Series 192). Amsterdam: John Benjamins. 107–125. https://doi.org/10.1075/slcs.192.05bro

Brown, Earl Kjar. (2020). The effect of forms' ratio of conditioning on word-final /s/ voicing in Mexican Spanish. *Languages* 5(4):no. 61. https://doi.org/10.3390/languages5040061

Brown, Earl K., & Alba, Matthew C. (2017). The role of contextual frequency in the articulation of initial /f/ in Modern Spanish: The same effect as in the reduction of Latin /f/? *Language Variation and Change* 29(1):57–78. https://doi.org/10.1017/S0954394517000059

Brown, Esther L. (2004). The reduction of syllable initial /s/ in the Spanish of New Mexico and southern Colorado: A usage-based approach. Doctoral dissertation, University of New Mexico.

Brown, Esther L. (2015). The role of discourse context frequency in phonological variation: A usage-based approach to bilingual speech production. *International Journal of Bilingualism* 19(4):387–406.

Brown, Esther L. (2018). Cumulative exposure to phonetic reducing environments marks the lexicon. In K. A. Smith & D. Nordquist (eds.), *Functionalist and usage-based approaches to the study of language: In honor of Joan L. Bybee* (Studies in Language Companion Series 192). Amsterdam: John Benjamins. 127–153.

Brown, Esther L., & Raymond, William. D. (2012). How discourse context shapes the lexicon: Explaining the distribution of Spanish f-/h words. *Diachronica* 29(2):139–161. https://doi.org/10.1075/dia.29.2.02bro

Brown, Esther L., Raymond, William D., Brown, Earl Kjar, & File-Muriel, Richard J. (2021). Lexically specific accumulation in memory of word and segment speech rates. *Corpus Linguistics and Linguistic Theory* 17(3):625–651. (doi:10.1515/cllt-2020-0016)

Brysbaert, Marc, & Diependaele, Kevin. (2013). Dealing with zero word frequencies: A review of the existing rules of thumb and a suggestion for an evidence-based choice. *Behavior Research Methods* 45(2):422–430. https://doi.org/10.3758/s13428-012-0270-5

Bybee, Joan. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14:261–290. https://doi.org/10.1017/S0954394502143018

Bybee, Joan. (2006). From usage to grammar: The mind's response to repetition. *Language* 82(4):711–733.

Bybee, Joan. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Bybee, Joan, & Napoleão De Souza, Ricardo. (2019). Vowel duration in English adjectives in attributive and predicative constructions. *Language and Cognition* 11(4):555–581. https://doi.org/10.1017/langcog.2019.32

Bybee, Joan, & Napoleão De Souza, Ricardo. (2021). The role of frequency and predictability in the formation of multi-word expressions. In A. Trklja & Ł. Grabowski (eds.), *Formulaic language: Theories and methods*. Berlin: Language Sciences Press. 3–29.

Cohen, Jacob. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen Priva, Uriel. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition* 160:27–34. https://doi.org/10.1016/j.cognition.2016.12.002

Cohen Priva, Uriel, Edelist, Lee, & Gleason, Emily. (2017). Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *The Journal of the Acoustical Society of America* 141 (5):2989–2996. https://doi.org/10.1121/1.4982199

Du Bois, John W., Chafe, Wallace L., Meyer, Charles, Thompson, Sandra A., & Martey, Nii. (2003). *Santa Barbara Corpus of Spoken American English* (Vol. 2). Philadelphia: Linguistic Data Consortium.

Eddington, David, & Channer, Caitlin. (2010). American English has go? A lo? Of glottal stops: Social diffusion and linguistic motivation. *American Speech* 85(3):338–351.

Egbert, Jesse, & Plonsky, Luke. (2020). Bootstrapping techniques. In M. Paquot & S. Gries (eds.), *A practical handbook of corpus linguistics*. Cham, Switzerland: Springer International Publishing. 593–610. (doi:10.1007/978-3-030-46216-1_24)

Erman, Britt, & Warren, Beatrice. (2000). The idiom principle and the open choice principle. *Text* 20(1):29–62.

Fletcher, Janet. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver & F. E. Gibbon (eds.), *The handbook of phonetic sciences* (2nd ed.). Hoboken, NJ: Wiley. 521–602.

Forrest, Jon. (2017). The dynamic interaction between lexical and contextual frequency: A case study of (ING). *Language Variation and Change* 29(2):129–156. http://dx.doi.org/10.1017/S0954394517000072

Fox, John, & Weisberg, Sanford. (2019). *An R companion to applied regression* (3rd ed.). Thousand Oaks, CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Freud, Debora, Ezrati-Vinacour, Ruth, & Amir, Ofer. (2018). Speech rate adjustment of adults during conversation. *Journal of fluency disorders* 57:1–10.

Gahl, Susanne. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3):474–496.

Godfrey, John J., Holliman, Edward C., & McDaniel, Jane. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of the* 1992 *IEEE International Conference on Acoustics, Speech and Signal Processing - Volume* 1, 517–520.

Gregory, Michelle L., Raymond, William D., Bell, Alan, Fosler-Lussier, Eric, & Jurafsky, Daniel. (1999). The effects of collocational strength and contextual predictability in lexical production. *Proceedings of the* 35*th Annual Meeting of the Chicago Linguistic Society*. 151–166.

Gries, Stefan Th. (2013). *Statistics for linguistics with R: A practical introduction* (2nd ed.). Berlin: De Gruyter Mouton.

Jacewicz, Ewa, Fox, Robert Allen, & Wei, Lai. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America* 128(2):839–850. https://doi.org/10.1121/1.3459842

Jurafsky, Daniel, Bell, Alan, Gregory, Michelle, & Raymond, William D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins. 229–254.

Kaźmierski, Kamil. (2020). Prevocalic t-glottaling across word boundaries in Midland American English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1):13. https://doi.org/10.5334/labphon.271

Kleiman, Evan. (2021). EMAtools: data management tools for real-time monitoring/ecological momentary assessment data (0.1.4) [Computer software]. https://CRAN.R-project.org/package=EMAtools

Kuznetsova, Alexandra, Brockhoff, Per B., & Christensen, Rune H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13):1–26. https://doi.org/10.18637/jss.v082.i13

Labov, William. (1984). Field methods of the Project on Linguistic Change and Variation. In J. Baugh & J. Sherzer (eds.), *Language in use: Readings in sociolinguistics*. Englewood Cliffs, NJ: Prentice Hall. 28–53.

Lehiste, Ilse. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Levshina, Natalia. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Lison, Pierre, & Tiedemann, Jörg. (2016). OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. http://opus.nlpl.eu/OpenSubtitles-v2018.php.

Nash, John C. (2014). On best practice optimization methods in R. *Journal of Statistical Software* 60(2):1–14. https://doi.org/10.18637/jss.v060.i02

Nash, John C., & Varadhan, Ravi. (2011). Unifying optimization algorithms to aid software system users: Optimx for R. *Journal of Statistical Software* 43(9):1–14. https://doi.org/10.18637/jss.v043.i09

Pépiot, Erwan. (2014). Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. *Speech Prosody* 7. 305–309. https://shs.hal.science/halshs-00999332

Pierrehumbert, Janet B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins. 137–157.

Pierrehumbert, Janet B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (eds.), *Laboratory Phonology 7*. Berlin: Mouton de Gruyter. 101–139.

Pierrehumbert, Janet B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (eds.), *Probabilistic linguistics*. Cambridge, MA: The MIT Press. 177–228.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye corpus of conversational speech* (2nd ed.). Department of Psychology, Ohio State University. https://buckeyecorpus.osu.edu/

Pluymaekers, Mark, Ernestus, Mirjam, & Baayen, R. Harald. (2005a). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62(2–4):146–159. https://doi.org/10.1159/000090095

Pluymaekers, Mark, Ernestus, Mirjam, & Baayen, R. Harald. (2005b). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America* 118(4):2561–2569.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing (v. 4.2.2) [Computer software]. https://www.R-project.org/

Sawilowsky, Shlomo S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods* 8(2):597–599. https://doi.org/10.22237/jmasm/1257035100

Schneider, Ulrike. (2020). ∆P as a measure of collocation strength: Considerations based on analyses of hesitation placement in spontaneous speech. *Corpus Linguistics and Linguistic Theory* 16(2):249–274. https://doi.org/10.1515/cllt-2017-0036

Segalowitz, S. J., & Lane, K. C. (2000). Lexical access of function versus content words. *Brain and Language* 75(3):376–389. https://doi.org/10.1006/brln.2000.2361

Seyfarth, Scott. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133(1):140–155. https://doi.org/10.1016/j.cognition.2014.06.013

Seyfarth, Scott. (2018). *Classes and iterators for the Buckeye Corpus* (v. 1.3) [Computer software]. https://github.com/scjs/buckeye/

Sóskuthy, Márton, & Hay, Jennifer. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition* 166:298–313. https://doi.org/10.1016/j.cognition.2017.05.032

Tagliamonte, Sali. (2006). *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.

Van Borsel, John, & De Maesschalck, Dorothy. (2008). Speech rate in males, females, and male-to-female transsexuals. *Clinical Linguistics & Phonetics* 22(9):679–685. (doi:10.1080/02699200801976695)

Wedel, Andrew B. 2004. *Self-organization and categorical behavior in phonology*. Doctoral dissertation, University of California—Santa Cruz.

Wedel, Andrew B. (2006). Exemplar models, evolution and language change. *The Linguistic Review* 23(3):247–274. (doi:10.1515/TLR.2006.010)

Wynn, Camille J., Barrett, Tyson S., & Borrie, Stephanie A. 2022. Rhythm perception, speaking rate entrainment, and conversational quality: A mediated model. *Journal of Speech, Language and Hearing Research* 65(6):2187–2203.

Wynn, Camille J., Borrie, Stephanie A., & Sellers, Tyra P. (2018). Speech rate entrainment in children and adults with and without autism spectrum disorder. *American Journal of Speech-Language Pathology* 27(3):965–974. (doi:10.1044/2018_AJSLP-17-0134)