S morpheme durations in English: An experimental approach

Noah Mahood Macey
Advised by Jason Shaw

Submitted to the faculty of the Department of Linguistics in partial fulfillment of the requirements for the degree of Bachelor of Arts.

Department of Linguistics
Yale University
April 24th, 2019

1 abstract

English has six kinds of suffixal morphemes that have [s] as an allomorph (so-called S morphemes):

1. Clitic-is: The cat's on the roof.
2. Clitic-has: The cat's been on the roof.
3. Plural: There are two cats.
4. Possessive: The cat's pajamas.
5. Plural-possessive: The two cats' hats.
6. 3rd-sg verbal suffix: The cat jumps.

Generative models of phonology, such as those following in Chomsky & Halle (1968), predict that these [s] segments should have the same phonetics independent of their morphemic identity. This prediction arises from a theoretical division between morphology and phonetics that prevents the two domains of language from talking to one another. Nevertheless, recent research has revealed that morphology and phonetics can, in fact, interact. One paper in this vein used linear mixed-effects modeling over a corpus of spontaneous speech to establish durational differences between these classes of S morphemes (Plag et al., 2017). This thesis presents the results of an experiment that tests whether these phonetic differences generalize to cases where the suffixes attach to nonce words. The results suggest that durational differences are not maintained in novel contexts, and thus that these phonetic differences conditioned by morphology do not generalize. It concludes with an explanation of the results through exemplar theory.

2 acknowledgments

First and foremost, this document would not and could not exist without the time and effort of my advisor Jason Shaw, who has shaped my approach to linguistic research since my first time doing it in his General Phonetics class and who has shaped this particular research through countably massive hours of his time and uncountably massive insights. A better thesis advisor I could not think of. This document is also indebted to Dolly Goldenberg, who volunteered several hours of her own time to get a forced aligner up-and-running on my computer and, when it failed to happen, literally gave me her computer for a full week while I ran the forced alignment of the data.

I cannot thank my friends enough, literally, because they were my subjects for this experiment and my IRB approval prevents me from supplying their names. Except for two—my roommate Vicky Blume, who kept me upbeat when it looked like this may never come to fruition, and my primary commiserator Sabrina Rostkowski, with whom I spent many late nights and early mornings working. Without her humor and support I would not have had the stamina to complete this process.

I owe much to all the linguistics teachers I've ever had—Jim Wood, Luke Lindeman, Chris Geissler, Josh Phillips, Sean Gleason, Sara Sanches-Alonso, Bob Frank, and Manu de Quadros, Raffaella Zanuttini, and Hadas Kotek—but I owe especially María Piñango, who has shaped my understanding of language, science, and clear thinking.

And of course I'd like to thank, my mother, father, and brother for not only allowing but actually *wanting* me to talk to them about this work. And thanks to my partners in crime, the senior seminar class: Jay, Jisu, Rose, Magda, Will, James, and Anelisa.

3 introduction to English S morphemes

Morphemes are the units of linguistic structure that represent pairs of sound and meaning (or, at least, sound and grammatical agreement). Morphemes have different functions—they can be lexical (i.e., change semantic content), like English *-ize*, or functional, like the past-tense *-ed*. The six S suffixes of English are lexical with the exception of the third-person singular agreement marker and the clitics representing *has* and *is*, and they are listed below (Plag et al., 2017):

1. Plural: There are twenty cat*s*.
2. Possessive (a.k.a. Genitive): The cat'*s* pajamas.
3. Plural-possessive: The twenty cat*s*' pajamas.
4. Third-person singular marker (henceforth 3rdsg): The cat jump*s* onto the sofa.
5. Clitic-*has*: The cat'*s* done it now.
6. Clitic-*is*: The cat'*s* a problem.

In the same way that phonemes can be realized as different allophones in different contexts (e.g., the change from a tap to a stop in the alternation between *atom* and *atomic*), a single morpheme can be realized as several allomorphs, also called exponents. The S suffixes of English, according to Bauer et al. (2013), appear as [-ɨz] following a sibilant (*i.e.*, /s, z, ʃ, d͡ʒ, t͡ʃ/), as [-s] following an voiceless consonant, and [-z] elsewhere. For the most part, these distributions apply to all of the S suffixes given above. Minor exceptions exist in the plurals, like the historically-based *leaf* $\sim$ *leaves* alternation (Bauer et al., 2013). Plag et al. (2017) also reports minor distributional differences for the possessive; however, overall there is no reason for systematic differences in how people pronounce the different flavors of S suffix, though the present investigation discovered one allomorphic variation (see section 9.1). They should appear as the same segments in the same contexts, and yet systematic difference is exactly what Plag et al. (2017) reports.

This leads us to a note on notation: I adopt Plag et al. (2017)'s convention of using capital-S as a stand-in for both /s/ and /z/ in word-final position. The term "S suffixes" refers to the whole set of six morphemes described above, while non-suffixal S refers to /s/ and /z/ sounds that terminate single morpheme words (such as the [s] in *lapse*).

4 introduction to homophony

Recent phonetic research (e.g., Gahl, 2008; Drager, 2011; Cho, 2001) has revealed subtle differences in the phonetic realizations of homophonous segments at the segment, morpheme, and word level—for example, differences in duration and vowel quality, or Cho (2001)'s observation that segment sequences across morphemes are more variable than the same segment sequences within morphemes. This paper investigates durational differences.

Many theories of speech production divvy words up into two parts: the phonological form, which is roughly the information about a word's sound, and the lemma, which is its meaning and syntactic category. This separation of the syntactico-semantic subsystem from the phonological subsystem makes it possible to measure a word's frequency in two different ways: the frequency of the phonological form and the frequency of the lemma (Gahl, 2008).

In general, more frequent words are shorter—but frequent by which measure? Gahl (2008) used English homophones to answer this question. Previous work (Jurafsky, 2003) suggested that function words do not exhibit shortening as a function of lemma frequency. Nevertheless, the linguistic machinery of function words can differ from that of lexical words, and Gahl (2008) used samples of homophone pairs with one high-frequency word and another low-frequency word to test whether shortening is a matter of lemma frequency or of phonological form. Since high-frequency homophones share their phonological forms with low-frequency homophones (e.g., *time* and *thyme*), it is possible to check whether or not the shortening they exhibit seems in line with their lemma frequencies or the frequency of their shared phonological form. Gahl (2008) uses regressions over 90,000 tokens in the CALLHOME corpus to do just that, and found that the phonological realizations of low-frequency homophones like *thyme* are significantly longer than those of their high-frequency counterparts, like *time*.

This body of research is twice important to the issue of S suffix duration in English. First, it demonstrates that duration can be based on lemmas. Second, it reveals that lemma frequency affects different swaths of the lexicon in different ways: function words do not seem to shorten based on lemma frequency, but lexical items do. In this paper, we are concerned with a third category of linguistic structure: suffixes below the word level. Do they exhibit durational differences? And if so, do these durational differences pattern like those of the lexical words (more frequent lemmas being shorter) or in some other way?

5 observed differences in S morpheme duration

Plag et al. (2017) asks exactly this question about the S suffixes, and one of the goals of the present research is to confirm their findings in the laboratory. In English, Plag et al. (2017) show that the six word-final homophonous S suffixes indeed differ in duration, both compared to lexical word-final S (e.g., *lapse* v.s. *laps*), as well as compared to each other (e.g., *laps* v.s. *lap's*). The connection to frequency is murkier at this level: frequency is not a significant predictor of S suffix duration. Lowest in duration of all the S suffixes are clitic-*is* and clitic-*has*, and though there are differences within the plural and genitive suffixes, these are less robust and disappear with voicing.

Though these durational differences are statistically robust, they are on the scale of milliseconds and are the result of large corpus studies—Plag et al. (2017) used the Buckeye corpus and linear mixed-effects regressions to analyze the differences in S morpheme length (Pitt et al., 2007). Meanwhile, Walsh and Parker (1983) tested a similar question using an experimental paradigm and actually found that word-final, non-suffixal S was shorter than suffixal S. Though their methodology does not pass muster by today's standards (Tomaschek et al., 2018), contradictory data warrants a follow-up.

5.1 a detailed look at Plag et al. (2017)

Since the present work relies on Plag et al. (2017) for direction, it is worth examining the study more closely. Using tagged data from the Buckeye Corpus of Conversational Speech (Pitt et al., 2007), the researchers ran statistical tests over 199 words containing a non-suffixal S and just over twice that number containing suffixal S: 95 tokens of the plural, 100 of the 3rdsg, 88 of the possessive, 47 of clitic-*has*, 95 of clitic-*is*, and 23 of the plural-possessive, for a total of 448 tokens. The aim was to reveal whether or not there were durational differences between suffixal and word-final, non-suffixal S, as well as differences within the class of morphemic S.

The statistical model took the form of a linear mixed effects model, in which the relationship between variables is modeled in tandem with the possibility of random effects (for a more complete explanation, see section 10.3). They made use of the following covariates to duration: local speech rate, base duration, voicing, number of syllables, number of consonants immediately preceding S, frequency, neighborhood density, bigram frequency, previous mention, and following context. When voicing is removed as a variable (as it is in the experiment here), the following factors significantly predict duration. From strongest predictor to weakest: Following context >> Type of S >> Base duration >> Number of following consonants >> Syllables per second. For unvoiced realizations, possessive, plural, 3rd-sg marker, and plural-possessive suffixes were realized as significantly longer than the clitics, and this is the pattern this paper sets out to replicate in the soundbooth.

6 theoretical implications

In models of generative phonology (foundationally, Chomsky and Halle (1968), phonological rules determine which allomorph of an affix appears in the environment. After this determination, the proper allomorph is stitched onto the base, and a process called bracket erasure eliminates the boundary between base and affix for the rest of phonological processing. In such a model, morphological boundaries should be invisible at the articulatory stages of processing. Studies like Cho (2001); Plag et al. (2017), and Gahl (2008) suggest that morphology can influence phonetics. The possibility that S suffix durations differ based on the kind of suffix is important to determine the exact nature of this interface: the roles of prosody and articulation, dialectical variation, and the like. It is also a possible locus of generativity within phonology and phonetics, as predicted by Jackendoff (1997).

7 critiques of corpus work

Though Walsh and Parker (1983)'s methodology and statistics are less rigorous than those of Plag et al. (2017), concerns about corpus based methodologies have been raised recently. Foulkes et al. (2018) encourage us to question whether or not the minute differences that corpus studies report are realized in everyday speech acts, or if they can only be seen over a large dataset. The answer to this question is relevant to determining the source of these variations, their effects on language change, and their implications for language architecture. Foulkes et al. (2018) highlight four problems with corpora—intra and inter corpus variability, resolution, and statistical robustness. Variability occurs within corpora because of the way they are collected, and across corpora because of the medium and circumstances of recording. Likewise, the forced-alignment programs used for annotation do not always work properly, and can be thrown for a loop by low-resolution recordings. This is especially relevant to work that relies on small durational differences. In summation, the general assumption that statistics will save you from noise in the data is not necessarily true when the data consists entirely of noise.

Other confounds that might affect variation are linguistic factors not parsable in the speech signal. Prosody is particularly unaccounted for, and much of the variability attributed to frequency and information content might be better attributed to speaker identity.

Another issue with statistical methods is the failure of coefficients and constants to correspond with real-life interpretations. Rarely do phoneticians map the statistics to the speech signal, and Plag et al. (2017) falls into this trap—it is well and good to find that the type of S suffix predicts the duration of the segment [s], but by how much does it change between forms? If the differences are minute, it is possible that the information is not accessible to the listener.

Foulkes et al. (2018) suggest that researchers do laboratory studies of corpus effects, and in light of these vagaries, it is worthwhile to attempt laboratory confirmation of Plag et al. (2017). Their results do help discriminate between theories of speech, but they require experimental bolstering in order to be taken seriously.

8 methodology

8.1 materials

Whereas Plag et al. (2017) relied on the linear mixed-effects model to “control” for various confounds on S-suffix duration, the experimental materials below are premised on controlling for these confounds experimentally. The stimuli consist of nonce words and sentential frames designed to elicit all the various types of S-suffixes on a single stem in similar contexts. In each of the following sections, I explain the design of these stimuli in terms of the covariates they hold constant.

8.1.1 nonce words

In total there were sixteen nonce words. They are shown below, orthographically and phonemically, followed by an explanation of why and how they were chosen.

1. bip /bɪp/
2. bimp /bɪmp/
3. vop /vɑp/
4. vomp /vɑmp/
5. dreep /dɹip/
6. dreemp /dɹimp/
7. sfop /sfɑp/
8. sfomp /sfɑmp/
9. flep /flɛp/
10. flemp /flɛmp/
11. floop /flup/
12. floomp /flump/
13. sklup /sklʌp/
14. sklump /sklʌmp/
15. strop /stɹɑp/
16. stromp /stɹɑmp/

The words are monosyllabic since there is evidence from English (Kim and Cole, 2005) and non-English (Nooteboom, 1972; Lindblom, 1963) languages that preceding syllables can affect duration. In addition, single-syllable stimuli made calculating the syllable-per-second speech rate easier. Each nonce word has /p/ as its final consonant for three reasons. First, /p/ is voiceless, so it will only take the voiceless allomorph [-s]. Plag et al. (2017) observed the greatest number of duration contrasts in this allomorph, so it is the one focused on in this experiment. If there is no durational difference in the voiceless case, then it is unlikely one would be seen in the voiced allomorphs, which are systematically shorter (Klatt, 1976).

Second, /p/ is made entirely with the lips; this allows the tongue blade maximal freedom to move into the position for /s/. Another lingual sound like /t/ might have created gestural interference since the tongue can only be in one place at a time.

Third, /p/ is a plosive, meaning that, in a waveform for the sentence, it appears as silence. The transition from silence to frication makes the boundary between /p/ and /s/ extremely easy to parse in a spectrogram, increasing the ease of visual confirmation of the boundary.

The nonce-words come in pairs based on the coda; for each token with a simple coda of the form /Vp/, there is a token with a complex coda of form /Vmp/. These pairs were constructed because the number of consonants preceding the S suffix was found to be a significant predictor of duration in Plag et al. (2017)'s model. Such a finding is in line with previous work showing that consonants in clusters shorten (Klatt, 1976).

In addition, the onsets vary in number of consonants and consonant type. I did not have access to neighborhood density data for these nonce words; however, neighborhood density was not found to be a predictor of S-duration in Plag et al. (2017), so it is unlikely that it would have an effect in an experimental context.

8.1.2 sentential frames

In total there were twelve sentential frames, shown below orthographically. An underscore represents the location of the nonce word, and the number in parentheses is the syllable count. Explanation follows.

1. The two ______s run together in the mornings. (11)
2. The ______ ’s run to work goes by a park. (9)
3. The two ______s’ run to work goes by a park. (10)
4. He ______s Rover the dog once a week. (9)
5. The ______’s run a marathon before. (9)
6. The ______’s running a marathon tomorrow. (11)
7. The two ______s key cars for fun sometimes. (9)
8. The ______’s key witness failed to appear in court. (11)
9. The two ______s’ key witnesses failed to appear in court. (13)
10. He ______s key donors in order to flatter them. (12)
11. The ______’s keyed plenty of cars in its day. (10)
12. The ______’s keying cars out front as we speak. (10)

As the syllable counts show, all sentences are roughly the same length. There are syntactic and semantic cues for the type of S: for instance, the word *two* preceding the plurals. Additionally, English orthographic conventions enable distinction between the kinds of S-suffixes. For more information about how I ensured that the participants understood which type of S-suffix was attached in each sentence, see section 8.3.

In Plag et al. (2017)’s models, the segments immediately following the S-suffix were the strongest predictor of /s/ duration. A following approximant was associated with the longest durations, and a following stop was associated with the shortest ones, with affricates, fricatives, nasals, and vowels occupying the middle of the distribution. In the sentences above, the first syllable following the S-suffix is always either *run, rover*, or *key* (the stop case). Thus, the sentences are designed to control for and verify this condition of the model.

8.2 participants

Data came from ten Yale College undergraduates ages 21 and 22, all native speakers of English from various locations in the United States. Six were women; four were men.

8.3 procedure

The recordings were taken in a sound booth using a logitech microphone. The participants were given sixteen shuffled sheets of paper, one for every nonce word, with the above sentences printed on them (in the same order they are presented above). I instructed each participant to begin by reading the sentences to her or himself and to understand how the word was being used (as a noun or verb). I then asked them to say the word in the frame, "This is a ______." Then, they proceeded down the list. If any garden path effects occurred, they were asked to read the sentence again until they understood it, and the token was rerecorded.

Participants began the actual sentence elicitations by reading the sentence out loud. After finishing, they looked up, away from the page, and repeated the sentence twice—they were instructed to do this as if in a conversation. The recordings were taken this way in light of criticism (e.g., Plag et al., 2017; Gahl et al., 2012) of methodologies in which stimuli are read out loud, since reading out loud tends to occur at a more regular pace than spontaneous speech. This regular pace can erase the effects of lexical frequency on duration, so it is controlled for here.

8.4 data processing

The initial recordings were taken into Audacity (Audacity Team, 2019), where the two conversational samples for each sentence were clipped into a `.wav` file. TextGrids were created in Praat (Boersma and Weenink, 2019) for each `.wav` file using the University of Pennsylvania forced aligner (Yuang and Lieberman, 2008). Since the aligner required files of sampling frequency of 11,025HZ and the sampling rate of the recordings was at a higher resolution (441,00HZ), running the files through Lennes (2017)'s Praat script `change_sample_rate_of_sound_files.praat` was necessary to perform the alignment. In order to obtain final labeled `.TextGrid` files, I manually inspected the aligner-generated `.TextGrid`s against the higher-resolution files, and corrected any inconsistencies in the sentence, nonce-word, and S suffix labels. The beginning of the S suffixes was taken to be the onset of aperiodic energy, including any /p/ burst—the burst requires a spread glottis, indicating that the /s/ gesture has already begun. The offset of the /s/ into /k/ was taken to be the beginning of silence, which occurs during the velar closure. The offset into /ɹ/ was taken as the beginning of voicing, which is characteristic of the approximant.

From the labeled intervals, I extracted the durations of each segment and word using the `duration_logger.praat` script (Crosswhite, 2009) and imported them into R (R Core Team, 2017), where all statistical analysis took place.

9 results

9.1 data exploration

Before we embark on our tour of the data, we should take stock of what data we have. As the experiment was designed, there were ten speakers, sixteen nonce words, twelve sentential frames, and two recitations of each nonce word-sentential frame combination, which should give a total of (10)(16)(12)(2) = 3,840 measurements.

Life being different from experimental plan, this is not the real number of measurements.

Subject no.5 was unable to pronounce *sfop* or *sfomp*, and the data for her tokens of *vomp* were corrupted during the Audacity saving process. Thus, we begin data exploration with (9)(16)(12)(2) + (13)(12)(2) = 3,768 tokens.

In the first place, we look to the differences in how individual speakers pronounce word-final S suffixes, as shown in Figure 1. [1] The subjects show considerable variability in

[1]Note that Figure 1 consists of boxplots; all boxplots in this paper are marked by a horizontal bar at the median, and the box shows the middle 50% of the data points. The lines extend to the full range of the data

the length of their pronunciation of S, and cursory inspection marks speakers 1 and 5 as pronouncing both plural and plural-possessive S longer than the other types of suffixes.

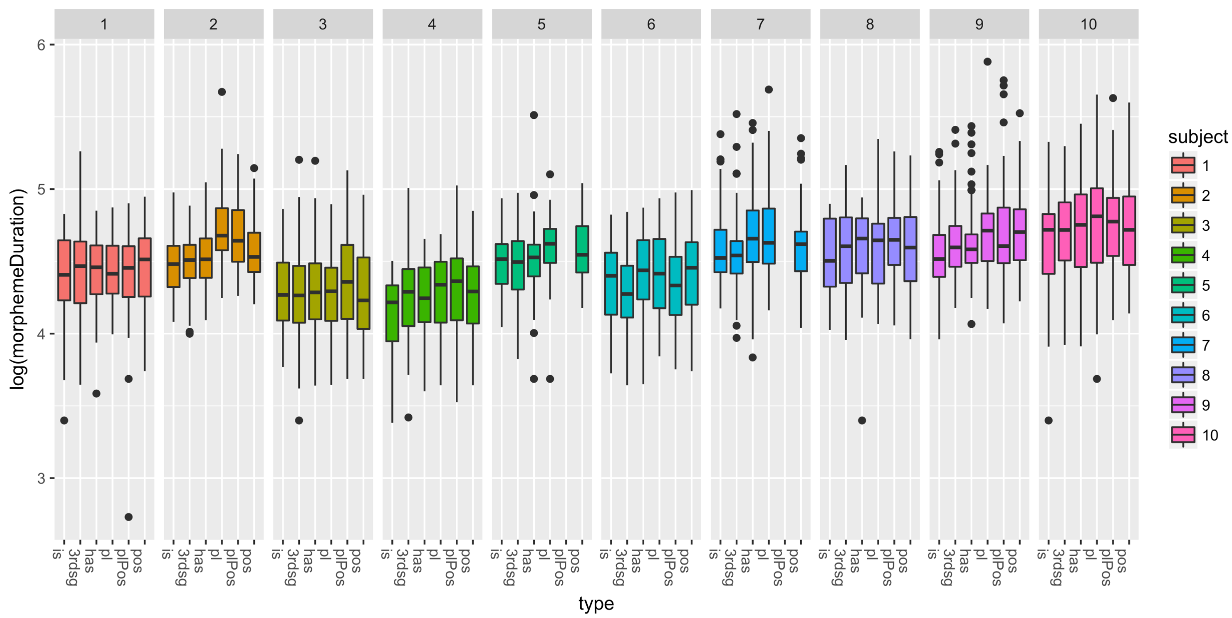

Figure 1: The x-axes show the type of S suffix, and the plot is tessellated by speaker. The y-axis shows the natural logarithm of the suffix duration measured in milliseconds (the logarithm is unitless). Explanations of the log-transform and the missing plural-possessive tokens for subjects no.5 and no.7 follow later in this section.

Next, we survey the suffix durations over the words: Did any nonce words prompt noticeably longer S suffixes? Figure 2 suggests no. All of the medians are between four and five. Some nonce words seem generally longer than others (e.g., *bips* v.s. *flemps*); however, no systematic structure of the words (number of segments, length of coda or onset, etc.) seems to condition this variation.

except for suspected outliers, which are shown as individual points.

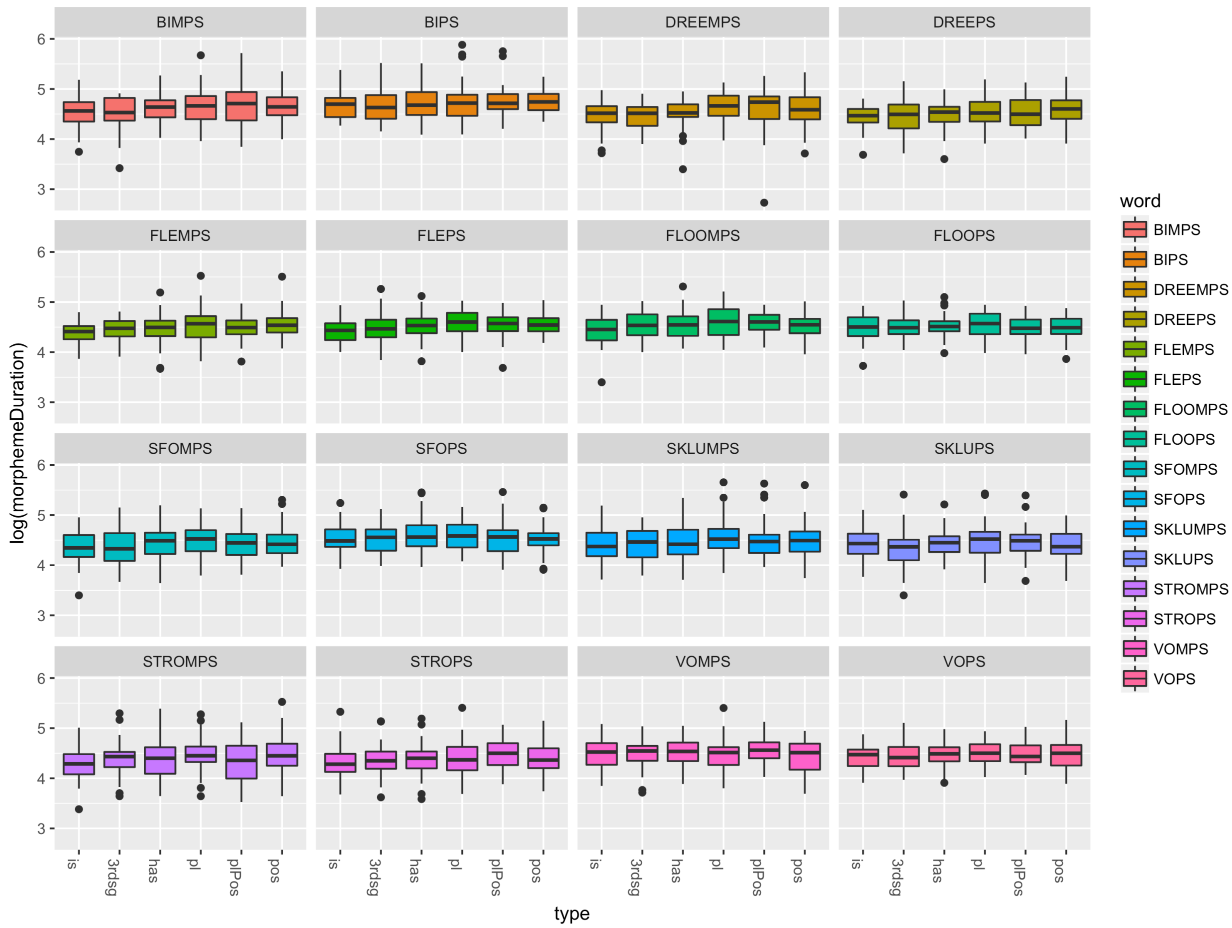

Figure 2: The x-axes show the suffix type in question, and the y-axis gives the logarithm of the duration (in milliseconds) of the associated word. All sixteen nonce words are present, tessellated in alphabetical order.

The mathematics of the linear mixed-effects model that we will use in future sections depends on the data involved being either divided into discrete categories or normally distributed. So as part of our preliminary screening, we check the shape of the data within each type of S suffix in figure 3. These distributions skew right and are not normal; however, this is typical of speech data and psychological experimental data more generally (Baayen and Milin, 2010). In many cases, the natural logarithm of the variable may be normal even if the variable is not. This is the case in Figure 4, which uses the natural logarithm of the durations on the y-axes rather than the durations themselves. The actual durations are easily recoverable by exponentiating the logarithm. This normality motivates the choice of log(suffixDuration) for the y-axes in figures 1 and 2.

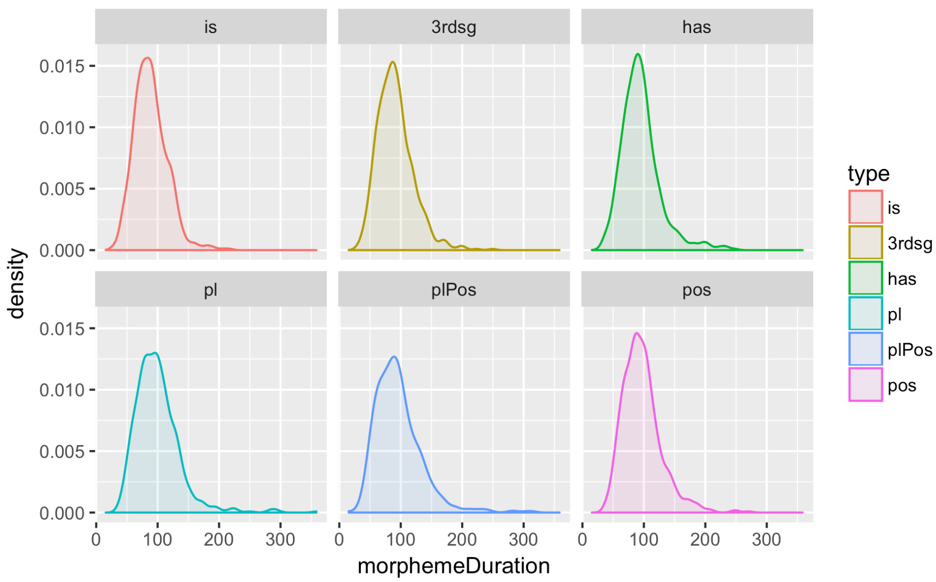

Figure 3: Kernel density plots for each S suffix type show the density of the suffix duration (in milliseconds) on the y-axes, and the duration of S suffix on the x-axes.

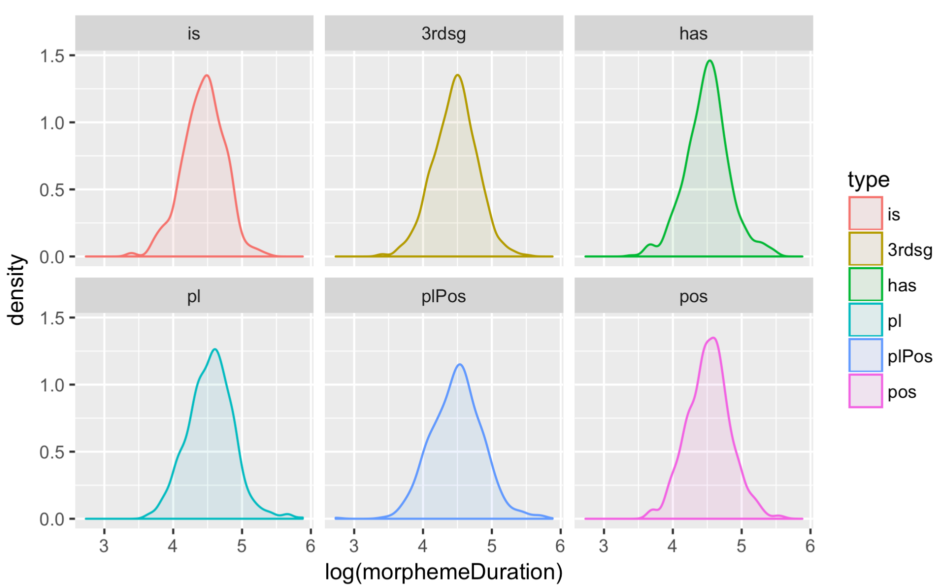

Figure 4: Kernel density plots show the density of the logarithm of the suffix duration (in milliseconds) on the y-axes, and the logarithm of the duration on the x-axes. Note the improved normality.

A major predictor of S suffix duration in Plag et al. (2017) is the duration of the *stem* (the word to which the S suffix is attached—e.g., "cat" in "cats"). So we must confirm normality in these as well. Figure 5 shows that the durations are not normal, and figure 6 shows that the logarithm of the stem durations is more normal, as in the case of the suffix durations.

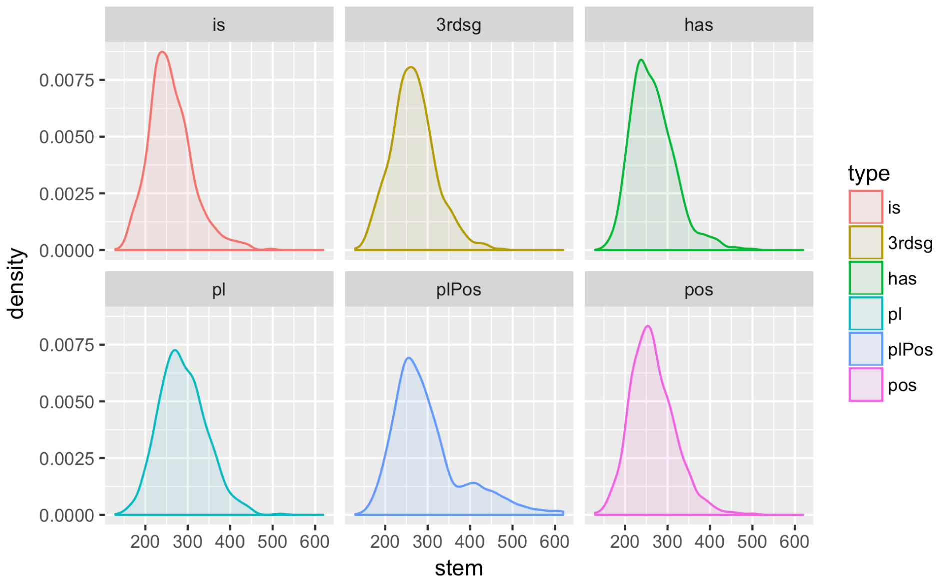

Figure 5: Kernel density plots showing the density of the duration (in milliseconds) of the stems on the y-axes, and the duration of the stem on the x-axes.

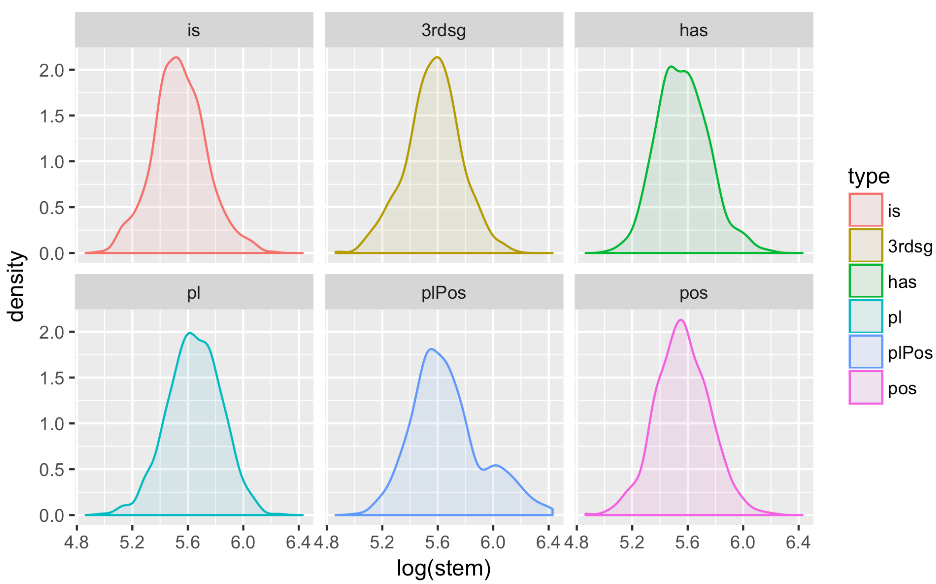

Figure 6: Kernel density plots showing the density of the log-transformed duration (in milliseconds) of the stems on the y-axes, and the log-transformed durations on the x-axes. Note the improved normality.

Nevertheless, the durations for the plural-possessive type of S suffix is bimodal in both figures 5 and 6, which reveals a non-normal distribution. The source of this abnormality can be traced to some unexpected dialectical variation among the subjects of the experiment. Two of them, subject no.5 and subject no.7, realized the plural possessive /s/ as [sɨz] in all contexts. For instance, they pronounced *bips'* as [bɪpsɨz], epenthesizing a vowel. This extra vowel adds length to the stem for a subset of the plural-possessive tokens. If we throw out these tokens, leaving us with (8)(16)(12)(2) + (16)(10)(2) + (13)(10)(2) = 3,652 data points, we see that the tokens of log(Stem duration) are now normal in all cases, as shown in figure 7.

Having removed these PLPOS tokens, it is worth checking to see if the log-transformed S suffix durations are still normal over each type. Figure 8 confirms that they are.

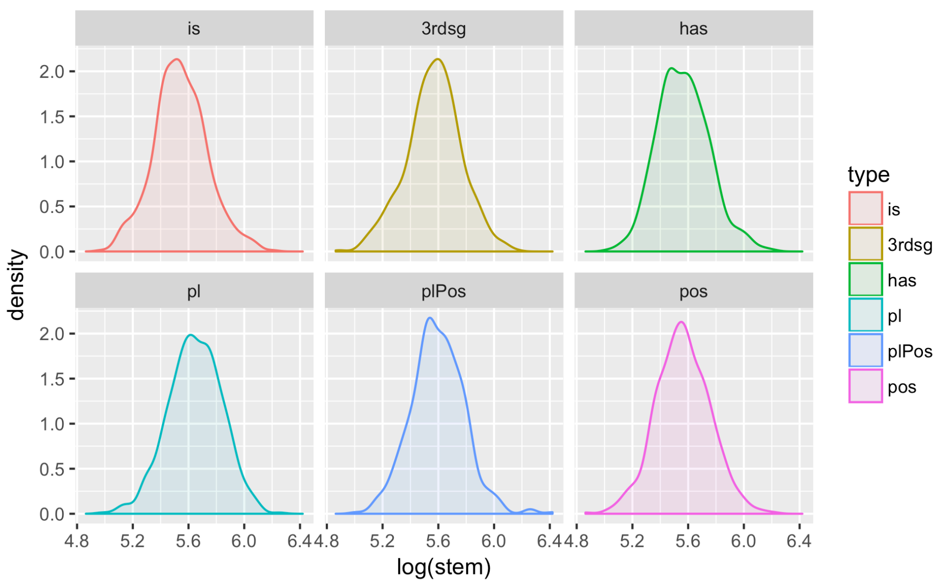

Figure 7: Kernel density plots showing the distribution of the logarithm of the stem durations on the y-axes (in milliseconds) and the logarithm of the stem durations on the x-axes for each type, with the plural-possessive tokens of subjects no.5 and no.7 removed due to dialectical variation.

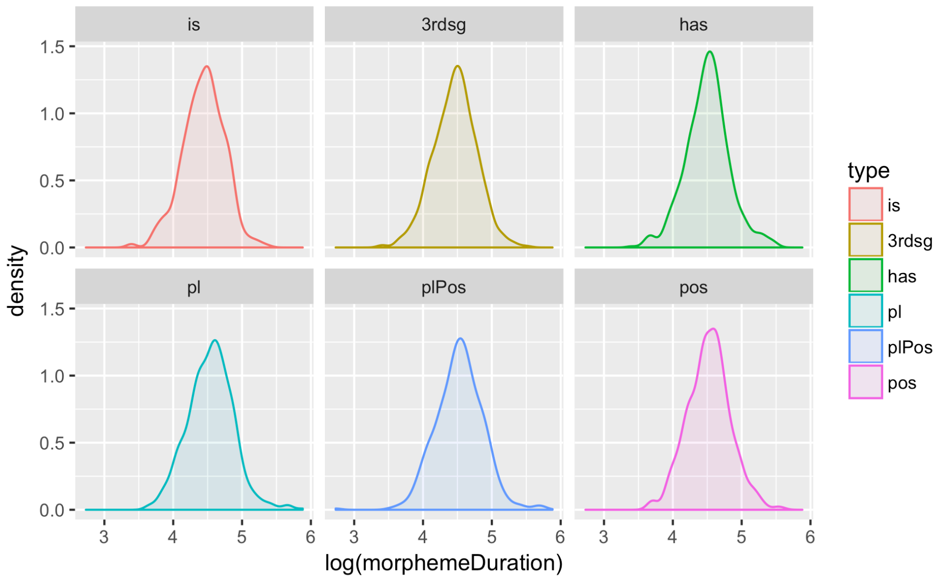

Figure 8: A kernel density plot showing the distribution of the logarithm of S suffix durations (in milliseconds) for each type, excluding plPos data from subjects no.5 and no.7.

The reader should note that from this point forward, all figures depicting the data use the restricted dataset of 3,652 points.

The data is now normalized, so a sanity check is in order. We have two measures, speech rate (syllables/second) and log(Stem) (unitless) that should, all else equal, correlate in specific ways with our data. In order to ensure that the collected tokens for suffix duration behave as they should, we can check their correlations with these two measures. Assuming that the correlations pan out the way we expect them to, then these variables will make for important fixed effects in the model we build in later sections.

There should be a negative correlation between speech rate and suffix duration because as people speak faster, they pronounce individual sounds for less time. This negative correlation is what we see in figure 9, and it is also found in Plag et al. (2017).

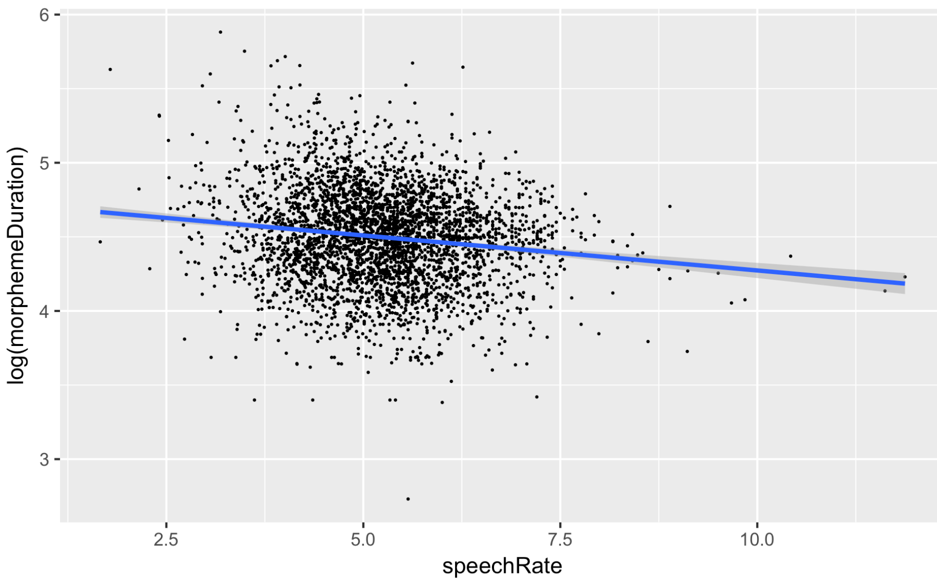

Figure 9: A scatter plot showing the distribution of the logarithm of suffix duration (in milliseconds) on the y-axis and speech rate, in syllables-per-second, on the x-axis. A best-fit line is shown to demonstrate the negative correlation.

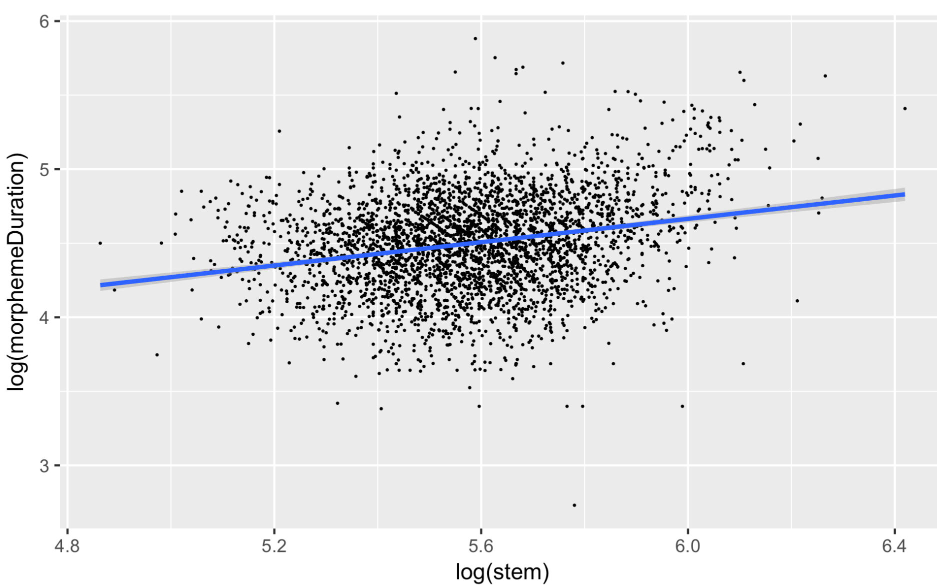

Figure 10: A scatter plot showing the distribution of the logarithm of suffix duration (in milliseconds) on the y-axis and the logarithm of stem duration (in milliseconds) on the x-axis. A best-fit line demonstrates the positive correlation.

Stem duration (and thus its normally-distributed proxy, the logarithm of stem duration) should correlate positively with suffix duration: if every other sound in a word is short, then the word-final sound, even if it is a separate suffix, should be short as well. This behavior is shown in figure 10.

Thus, the suffix durations behave as we expected them to in relation to speech rate and stem duration, which is a vote of confidence in the accuracy of our data. Both of these covariates will appear as fixed effects in our model.

Another indication that the experimental data does not reflect natural speech is a durational difference between the first and second elicited S suffix for a given word-sentence combination. If the second repetition of the suffix is significantly shorter than the first, then the articulators might be reciting automatically, rather than as they would in a normal speech act. For instance, if the S suffix in “The two dreemps run together in the mornings” is much longer in the first token than in the second, it is an indication of a different sort of motor control over the articulators. Figure 11 shows no apparent difference between the first and second elicitation of the S suffix for a given sentence.

Figure 11: A boxplot showing the logarithm of the duration (in milliseconds) of S suffixes on the y-axis. The x-axis is divided into categories for whether the suffix was the first or second repetition of the nonce word—the recitation order.

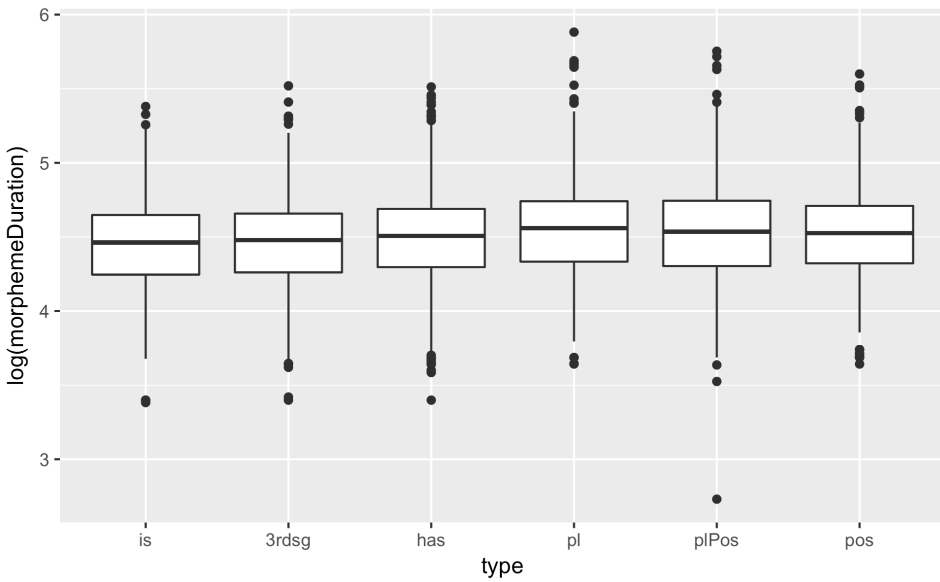

Figure 12: The y-axis represents the logarithm of the suffix durations (in milliseconds) and the x-axis shows the types of suffixes. Note that the medians of the plural, plural-possessive, and possessive suffixes are noticeably higher than the 3rdsg, clitic-is, or clitic-has suffixes.

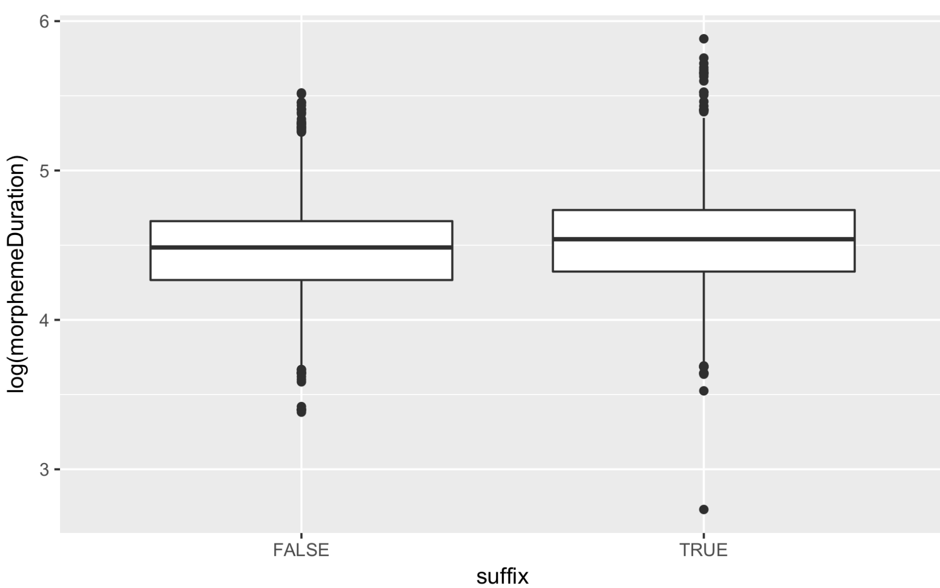

Figure 13: The y-axis represents the logarithm of the suffix durations (in milliseconds) and the x-axis shows the types of suffixes where plural, plural-possessive, and possessive have been collapsed into one, vaguely-named category SUFFIX.

We look at whether or not the suffix durations differ based on the type of suffix. Figure 12 shows higher medians and distributions for the plural, plural-possessive, and possessive suffixes than for the others, which is consistent with Plag et al. (2017). If we collapse these three suffixes into one class, as in figure 13, the difference becomes even clearer.

The strongest predictors of suffix length reported in Plag et al. (2017) were the consonant immediately following the S suffix. Approximants, like /l/ or /ɹ/, lengthened suffix durations while stops shortened them. This difference is natural from an articulatory standpoint: the occlusion of the airway in a stop causes an immediate cessation of the airflow necessary to sustain [s], while approximants do not greatly occlude airflow. The sentential stimuli tested this prediction; half of the S suffixes were positioned preceding a word beginning with the approximant /ɹ/ and half preceded /k/. Figure 14 shows that the S suffixes preceding the approximant were indeed much longer than those preceding the stop, bearing out the effect found in Plag et al. (2017). One question is whether or not the different types of S suffixes are affected by the following consonant in different ways. For example, does the plural suffix lengthen before /ɹ/, but not the possessive suffix? Figure 15 compares how the type of the suffix interacts with the following consonant to produce duration differences, and it does not show any major effects—perhaps marginally, the plural, plural-possessive, and possessive suffixes lengthen more in front of the approximant than do clitic-has, clitic-is, and the 3rdsg suffix. We will explore the possibility of this interaction more in future sections.

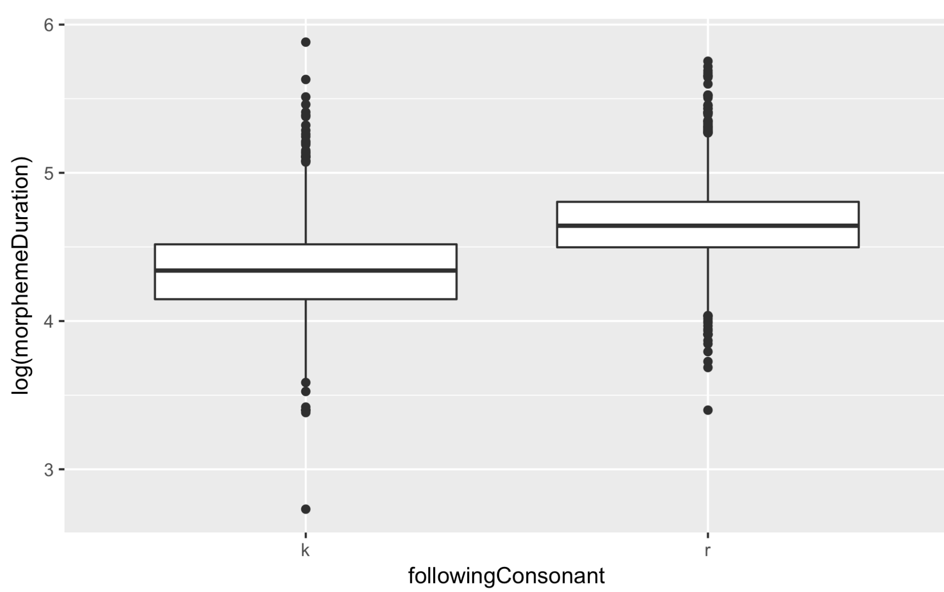

Figure 14: This boxplot, with the log-transformed suffix duration (in milliseconds) on the y-axis and the following consonant on the x-axis, shows that the [s] segments preceding the approximant were much longer than those preceding the stop.

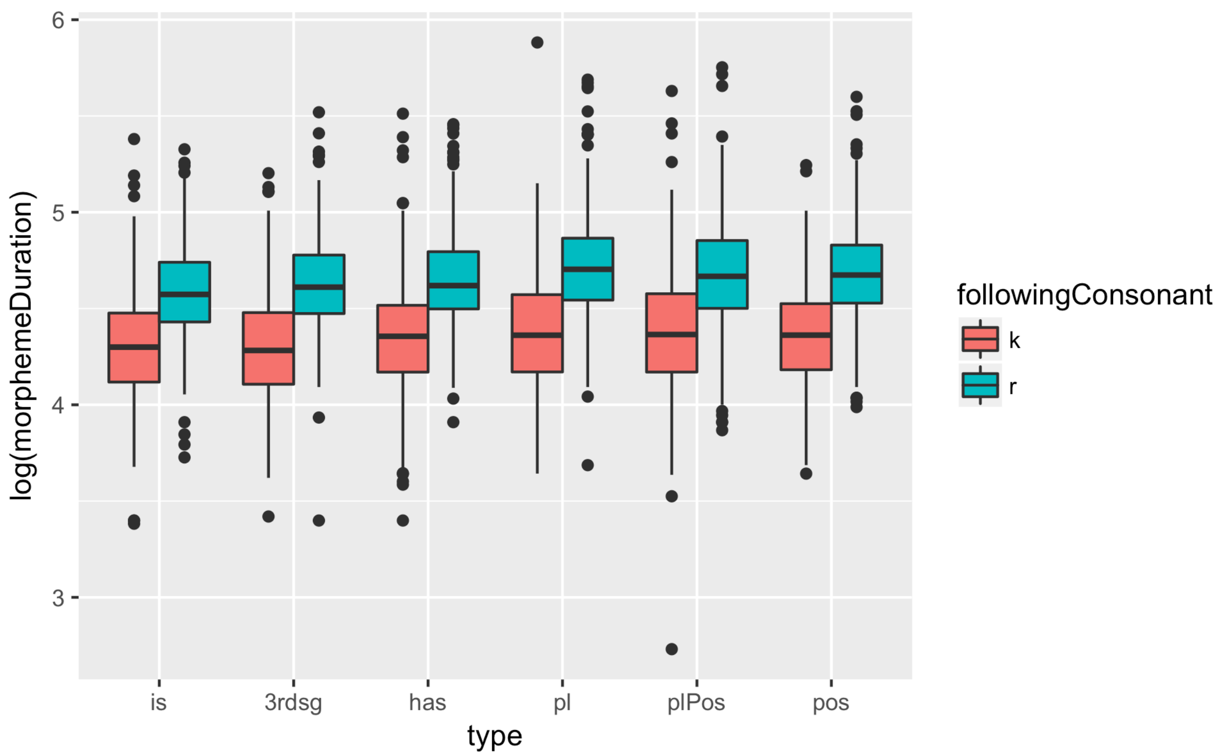

Figure 15: The y-axis displays the log-transformed suffix duration in milliseconds, and the x-axis displays the suffix type. The following consonant, whether /k/ or /r/, is shown via the color of the box. Note that all suffix types lengthen in front of the approximant, but that plural, plural-possessive, and possessive appear to lengthen slightly more.

Plag et al. (2017) found that the number of consonants preceding the S suffix significantly predicted the suffix's length, and there is cross linguistic evidence that increasing the number of segments in a given morphological unit leads to compression of the segments. That is, the more sounds in a phrase, the shorter each sound becomes (Klatt, 1976; Nooteboom, 1972; Lindblom, 1963), thus it was included in experiment here, where the nonce words ended in either /p/ or an /mp/ cluster. Figure 16 shows that this condition did not have an apparent effect on S suffix duration, which is out of step with the previous work.

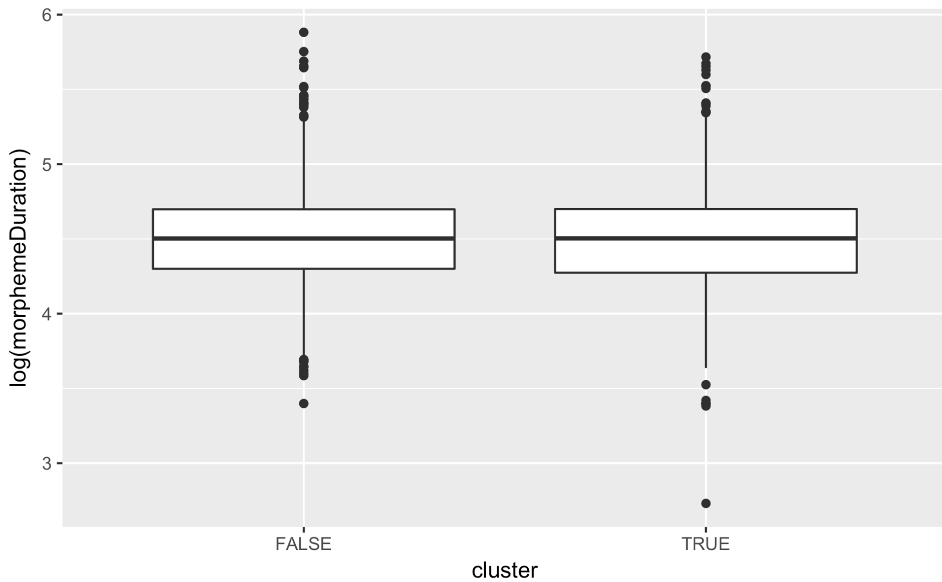

Figure 16: On the y-axis, we have the log-transformed S suffix duration (in milliseconds), and on the x-axis we have two conditions—when CLUSTER is TRUE, the suffix was attached to a nonce word ending in /mp/, and when CLUSTER is FALSE, the nonce word ended in a singleton /p/. There is no visible difference between the two conditions, indicating no compression took place, or that any compression was minor.

Another place to look for compression effects is the relationship between S suffix length and stem length, shown in figure 10. In figure 10, the S suffix compresses along with the stem: as the stem shortens, so does the S suffix. However, since the entire conceit of this paper is to explore these S suffixes as a non-homogenous group, it makes sense to wonder whether or not they are all equally vulnerable to compression effects. Figure 17 explores this question. Steeper slopes indicate that the suffix compresses more, that a shortening of the stem does leads to a shortening of the suffix. Steeper slopes appear for the plural and plural-possesive suffixes, while the clitics are quite level.

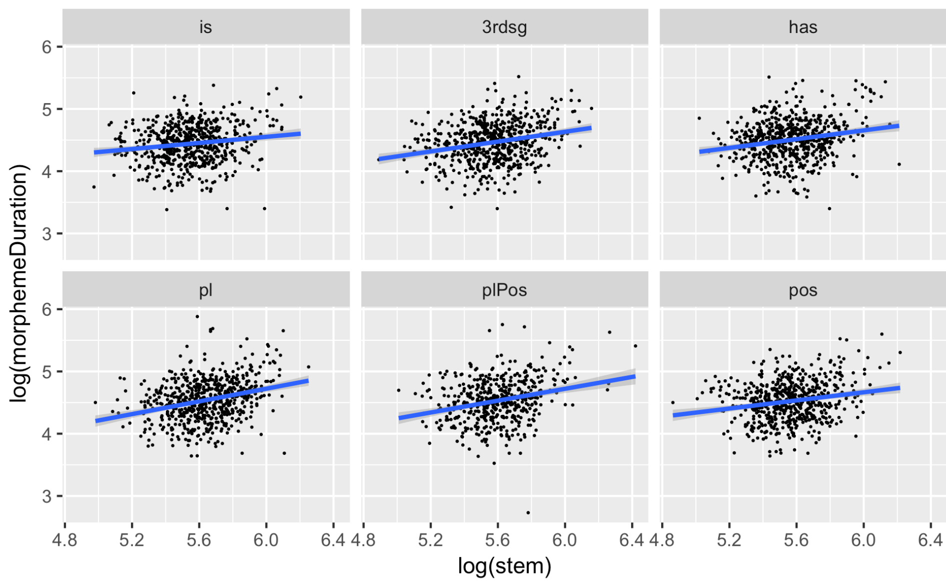

Figure 17: On the y-axes, we have the logarithm of the suffix duration in milliseconds. On the x-axes, we have the logarithm of the stem duration in milliseconds. There is one scatter plot per type of S suffix, with a line of best fit displayed to show a positive, negative, or lack of correlation. The flatter slopes indicate less compression of the S suffix along with the stem (see the preceding paragraph for a full explanation).

9.2 summary of results

The above visualizations alert us to a number of trends in the data. The duration data is right skewed—typical of this type of data—but the log-transform is not. Speakers no.5 and no.7 caused non-normality in the data for the plural-possesive suffixes due to dialectical variation, so their tokens for those suffixes were removed from the dataset.

Stem duration and speech rate both correlate with S suffix durations, as is expected, and the S suffixes of plural, plural-possessive, and possessive types appear to have slightly longer durations. Like Plag et al. (2017), the type of consonant following the S suffix is a powerful predictor of duration; unlike Plag et al. (2017), the number of consonants preceding the S suffix is not. These consonant and stem duration variables seem to interact with the type of suffix, as shown in figures 15 and 17: the different S suffixes respond differently to changes in predictors, bolstering the hypothesis that they are not a homogenous class when it comes to phonetic behavior.

All of these trends amount to hypotheses; the visual inspection of the data cannot provide statistical significance. In the following section, linear mixed-effects modeling is used to test these observations for significance, with the ultimate goal of determining how homogenous the pronunciation of the S suffixes are in different contexts.

10 analysis

The next sections are laid out as follows: In section 10.1, I explain how the mathematical model we will construct will provide us with an answer to our question—whether the English S suffixes differ systematically in length—and how we will go about constructing it. We test for patterns and problems in the data by constructing various graphs, we correct the problems, and we construct the actual model and determine whether or not the different S suffixes were produced with different durations.

10.1 linear mixed-effects modeling

There are many reasons that the given sounds in a given word are pronounced the way they are: where the word is said, why it is said, who says it, how loudly, to whom, in a statement or question, etc. Plag et al. (2017) found that one of these factors, for the /s/ sounds at the end of words, is the type of S suffix. The team did so using a *linear mixed-effects model*, a kind of mathematical model in which a given phenomenon—here, the log-transformed duration in milliseconds of an S suffix—is described as the combination of random and fixed effects. These effects are variables that help predict the target variable (the thing we want to predict) that the researchers are modeling. For example, say our target variable is tomorrow's temperature in a given city. Two effects we would like to include in our model are the temperature today and the average temperature of tomorrow's date over the past ten years.

The usefulness of linear mixed-effects models comes from their division of effects into *random* and *fixed* (for a detailed and accessible guide to linear mixed-effects models, and linear models in general, please see Winter (2013), from which all of this information is adapted.) A fixed effect is a variable that we expect to determine the target variable in a specific, systematic way. For instance, for the issue at hand here, we expect that the speech rate of the subjects will systematically affect the duration of the S suffixes. The higher the speech rate, the shorter the suffix; this is what is meant by ”systematic.”

One reason to use linear mixed-effects models here is that they provide information about how good the fixed effects are at predicting the target variable. They tell us whether or not a given effect is involved in the target variable outcome. In our temperature predicting example, it would be possible to include day of the week as a fixed effect; however, the model will, through a lack of statistical significance, reveal this as a terrible predictor.

Random effects, by contrast, are variables we believe will affect the target variable, but it is more difficult to predict in what way the effect will display itself. For example, consider speaker identity. It makes complete sense that different speakers will pronounce words differently, but there is no easy *a priori* way to determine whether each speaker will pronounce a word faster or slower than any other. Colloquially, one could say that the speaker introduces a little bit of randomness to the model, and it is just these sorts of unpredictable, non-systematic variables that compose a model's random effects. More precisely, a random effect is sampled from the population and does not exhaust it. For instance, there are many more possible subjects and nonce words than those used in this experiment, but there are no other kinds of S suffixes, so the data exhausts the population of S suffixes but not of subjects and nonce words.

In sum, a linear mixed-effects model is a way of predicting the value of one target variable based on several fixed and random effects, which have a relationship with the target variable. They also provide information about whether or not a given effect is statistically significant with respect to the target variable. In our case, the target variable is the *duration of suffixal S*, and the crucial question is whether the *type of S* proves to be a statistically significant fixed effect. Before we test that we must add to the model all other fixed effects which we know affect S suffix duration. The type of S is only a significant new predictor if it makes a model based on the other predictors more accurate. We explored many possible predictive factors in section 9.1.

Before we begin choosing fixed effects, we must decide on random effects. As seen in the methodology section, experiments were highly controlled. Nevertheless, randomness enters our data in three main ways.

The first is through the subject identities. Each experimental subject adds a bit of randomness to the data and comes nowhere close to exhausting the population of English speakers. Adding SPEAKER as a random intercept (the standard way to add this sort of random effect) allows the model to account for how much variation is due to the subject rather than the S suffix type or any other fixed effect. To view the variation in S pronunciation by the ten subjects, see figure 1.

The next random effect is the effect of the individual nonce words. Each nonce's string of segments—really a string of muscle movements—cannot be said to interact with the S suffix durations in any given way, but it likely does since some motor sequences are easier and more fluid than others. Likewise, the experiment did not exhaust all nonce words.

The final random effect we include in the model is that of the sentence. Each nonce word cycled through all twelve sentential frames over the course of the experiment, so it is possible to tease out any random effect that the sentences may have imposed over the S suffix duration independently of the nonce words.

10.2 data preparation

One way to build linear mixed-effects models (which will be done in the next section), put forth by Baayen and Milin (2010), is to preliminarily trim a small number of outliers from the dataset before embarking on the analysis. Then, after the major fixed and random effects have been decided on, another small trimming of outliers may take place before the model is tested. The preliminary trim, in this case, targeted the most extreme 1% of tokens of the log-transformed S suffix durations (thirty-seven of the 3,652 data points remaining after removing the deviant plural possessive tokens). Figures 18 and 19 are probability plots of the data, indicating that the removal of these 37 outliers bring the dataset closer to normality. All the linear mixed-effects models here use this trimmed dataset of 3,615 tokens. See table 1 for a breakdown of how many tokens of each suffix type are used for modeling.

clitic-is	3rdsg	clitic-has	plural	plural-possessive	possessive
625	624	620	618	503	625

Table 1: This table displays the number of tokens of each S suffix type used in the linear mixed effects models below.

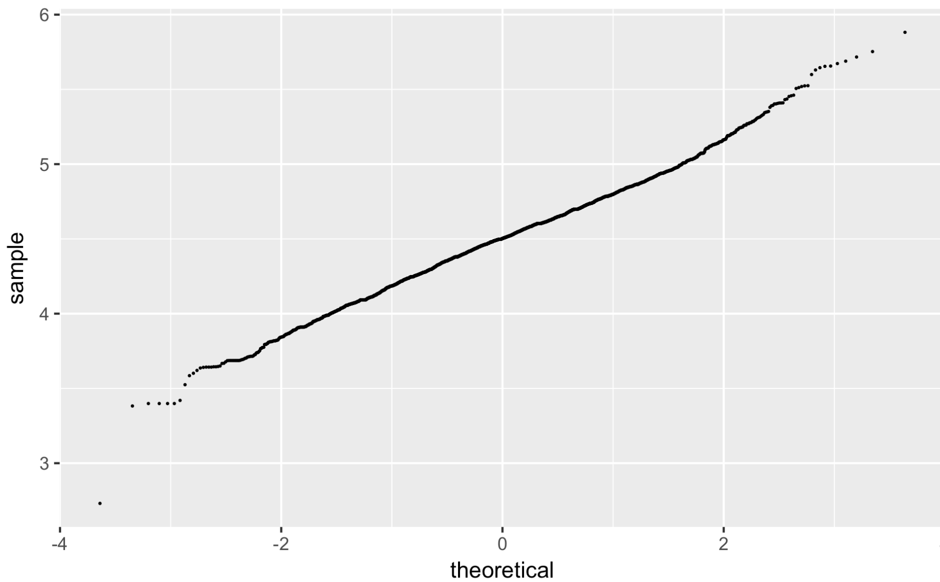

Figure 18: A probability plot, with all the log-transformed S suffix durations on the y-axis and theoretical quantiles on the x-axis, including the 1% outliers. A normally distributed data set will appear as approximately a 45°line through the origin. Note the deviancy from normality at the ends.

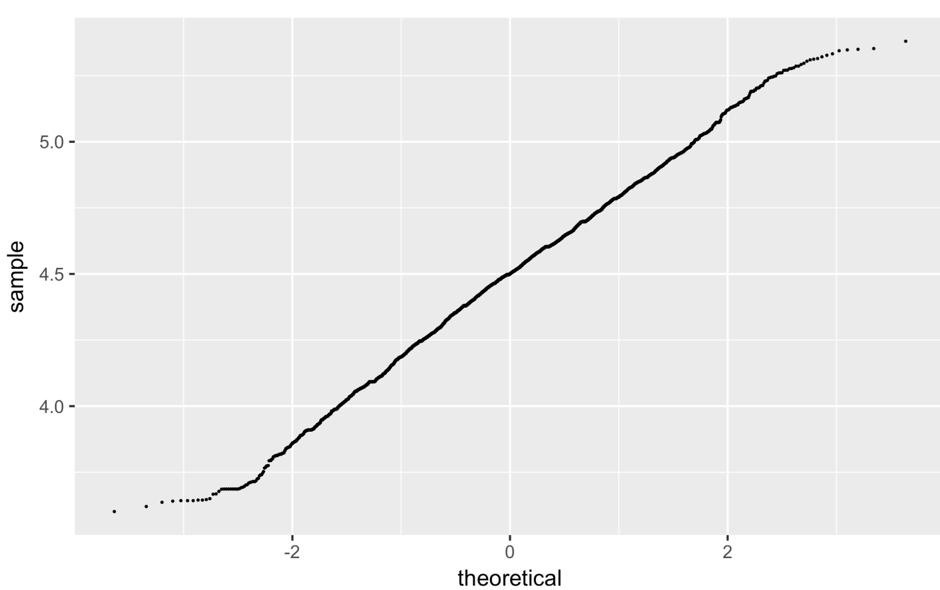

Figure 19: A probability plot, with the trimmed log-transformed S suffix durations on the y-axis and theoretical quantiles on the x-axis. A normally distributed data set will appear as approximately a 45°line through the origin. Note that this set, though still nonlinear at the ends, is more linear than the one in figure 18

10.3 model building

All linear mixed-effects models in this paper were created in R through the `lme4` package (R Core Team, 2017; Bates et al., 2015). Results of each model, and their comparison via ANOVA, can be found in the Appendix, with which the reader is encouraged to interact throughout this section. The initial model contains three random random effects as random intercepts: the subject, the word, and the sentential frame, with the log-transformed S suffix as the dependent variable. The only fixed effect was speech rate (SPEECHRATE), in syllables per second. From there, the following variables were added one at a time (random effects held constant), in the following order:

1. Log-transformed stem duration (LOG(STEM))
2. Following consonant—either /r/ or /k/ (FOLLCONS)
3. Consonant cluster in coda—did the nonce word end in /mp/? (CLUSTER)
4. Recitation order—was the suffix the first or second in the repetition of the frame sentence? (RECITATIONORDER)

After the residuals were tested for normality and non-heteroskedasticity (again, see Winter (2013)), the models were tested for significance via a likelihood ratio test, performed with pairwise ANOVAs. As an illustration of how this works, consider the model with both LOG(STEM) and SPEECHRATE. In order to determine whether or not this model was significantly better than the model based only off of SPEECHRATE, I ran an ANOVA comparing the two models, which returned three pieces of information: both models' AIC, both models' BIC, and a likelihood ratio. The AIC and BIC are two slightly different measures that penalize complexity and overfitting, and lower values for these criteria indicate a better model (Kuha, 2004). Higher values suggest that the model may be learning the specific data points rather than the trends underlying these points. The likelihood ratio is similar to a probability value, with statistical significance set at or below 0.05. So in the current example, the model using only SPEECHRATE had AIC = -1,378.5 and BIC = 1,341.3. The model using both SPEECHRATE and LOG(STEM) had AIC = -1,764.5 and BIC -1,721.2—both criteria were lower than the single-variable model. The likelihood ratio was 2.2×10^{-16}, far below 0.05. Therefore, we conclude that LOG(STEM) is a significant predictor of log-transformed S suffix duration, and we retain it in our model.

These three criteria—lower AIC, lower BIC, and likelihood ratio below 0.05—were met by the addition of log(Stem), followingConsonant, and recitationOrder, but not by cluster. Therefore, cluster was left out of the model. The resulting model of the log-transformed S suffix durations as a function of four fixed effects and three random effects is our baseline model; in order to justify any additional explanatory variable, it must improve on this baseline. Table 10.3 below gives the coefficients, t-values, and predicted unit change in milliseconds of the baseline model.

fixed effect	coefficient	t-value	unit change in ms
speechRate	-0.076	-13.058	-1
log(Stem)	0.510	19.146	2
followingConsonant = /r/	0.350	13.106	1
recitationOrder = second	-0.045	-6.305	-1

Table 2: Coefficients, t-values, and unit change in milliseconds (rounded to the nearest whole number) for the fixed effects of the baseline model. The reference value of followingConsonant = /k/ and the reference value of recitationOrder = first. The unit change in milliseconds refers to the expected increase or decrease in S suffix duration for a unit increase in a fixed effect. For categorical effects, it is the expected change in ms expected given the change in reference value.

Now, we compare what happens if we add the type of S suffix (type) to the model. If it meets the criteria above, then the model confirms the hypothesis that suffix type is a significant predictor of duration.

The addition of type to the model yields the model represented in table 10.3; however, the AIC of the model increases by 7.3, the BIC increases by 38.3, and the likelihood ratio is 0.748, so the model is not significantly better than the model without type.

FIXED EFFECTS	coefficient	t-value	unit change in ms
INTERCEPT	1.87	10.8	6
SPEECHRATE	-0.075	-12.9	-1
LOG(STEM)	0.51	19.3	2
FOLLOWINGCONSONANT = /r/	0.35	14.95	1
RECITATIONORDER = second	-0.045	-6.2	-1
TYPE = 3rdsg	-0.021	-0.52	-1
TYPE = has	0.017	0.42	1
TYPE = pl	-0.026	-0.63	-1
TYPE = plPos	0.027	0.67	1
TYPE = pos	0.021	0.52	1

Table 3: Coefficients, t-values, and unit change in milliseconds (rounded to the nearest whole number) for the fixed effects in the model including the baseline variables and the type of S suffix. The reference value of FOLLOWINGCONSONANT = /k/, the reference value of RECITATIONORDER = first, and the reference value of TYPE = "is."

Nevertheless, the type of suffix might still have significant *interactions* with other variables, and it may still be a significant predictor for individual speakers.

Indeed, when we look at the interactions between S suffix type and stem duration (the sort of relationship explored in figure 17), we find that the plural, plural possessive, and possessive suffixes become more strongly predictive than when they are divorced from the stem duration; however, compared to the baseline, this model does not pass muster—though the AIC of the model with the interaction is 0.3 lower, the BIC is 30.7 higher, and the likelihood ratio is 0.067.

One way to make statistics more meaningful is to reduce the degrees of freedom. Perhaps, since plural and plural possessive seem to pattern together in figures 15 and 17, the only salient morphological factor for S suffix duration is whether or not the suffix is plural. To that end, we consider two other models. The first, the ISPL model, has a fixed effect ISPL, either true (for the plural and plural possessive suffixes) or false (for all other suffixes). The second model, the interaction model, has a fixed effect for the interaction between ISPL and LOG(STEM). The ISPL model has higher AIC and BIC values than the baseline model and lacks statistical significance. The interaction model is statistically significant and has a lower AIC, but has a higher BIC. Compared to the ISPL model, the interaction model is marginally significant, with likelihood ratio 0.0046, a lower AIC, and a BIC that is higher by 0.3. Therefore, plurality alone and an in an interaction with the stem duration is not significant.

Breaking the data down by subject, ANOVAs comparing a subject's baseline model (i.e., a model with the fixed effects SPEECHRATE, LOG(STEM), FOLLOWINGCONSONANT, RECITATIONORDER and random intercepts for nonce word and sentential frame) to a model including TYPE yielded three significant likelihood ratios: 0.011 for subject 4, 0.0065 for subject 5, and 0.034 for subject 7. Nevertheless, none of these models including TYPE yielded lower BICs, so these results do not satisfy our criteria. Various data-pruning techniques—removal of more outliers, collapsing the types of S-suffixes so as to reduce the degrees of freedom—do not nudge the model close to statistical significance; thus, we must reject the proposition that the type of S suffix has a significant effect on S suffix duration in nonce words.

11 discussion

To recapitulate, none of the models we have just examined in the previous section replicate the findings with respect to morpheme type in Plag et al. (2017). In failing to do so, it confirms another paper, Foulkes et al. (2018), which expresses doubt that the phenomena observed in large scale corpus studies are always replicable experimentally. The null result presented here begets questions about the relationship between corpus work and experimentation. In the what follows, I remark on this relationship and the sorts of mental grammars that can contain both the present finding and that of Plag et al. (2017)—a grammar where S affix semantics affect /s/ duration in spontaneous speech, but not in novel contexts.

The finding here pertains only to the *generalizability* of the S morpheme duration findings in Plag et al. (2017) to nonce words, not the fact of the phenomenon in the wild. An experimental replication with extant words, while much more difficult to control, would be necessary to speak to whether or not that particular phenomenon could be found in the lab. Nevertheless, the fact of duration differences in S affixes exclusively with extant words and not in nonce words has implications for the structure of the language faculty.

One interpretation of the results in Plag et al. (2017) is that the different S affixes are specified down to the level of phonetic, rather than phonemic, detail. If this information were truly coded into the affix, then this experiment would have demonstrated these differences, which did not happen. Instead, some other pressure drives the affixes apart in the normal English lexicon. One candidate could be some covariate not included in the battery used by Plag et al. (2017), such as the context of the utterance, but a variable that would systematically affect S affixes without being folded into variables like speech rate. Indeed, the team was so thorough that it is hard to imagine a covariate they overlooked. More likely is that their data pool might have been too small, and the effects they saw were by chance rather than by grammar.

Another explanation comes to us from exemplar theory, a theory in which speakers sample from examplar banks linguistic categories (e.g., words, affixes, phonemes). The banks are experiential memories of past instances of these categories, rather than abstract representations (Winter and Wedel, 2016). The participants in this study have no exemplars of either the target words or the word-affix combinations, and this corresponds to a lack of difference in the S affix. In real words, then, the differences in S affixes could be the product of something like genetic drift, in which random variation in S affix duration when attached to various words replicates itself by entering the exemplar banks of the speakers. In this phonetic drift, the plural variant, say, may become longer simply because random lengthening replicated itself. *This lengthening does not occur in the actual abstract lexical item corresponding to the suffix*, rather, it occurs over the composed items. The results reflect a system in which commonly composed stem-suffix pairs develop exemplar banks of acoustic, motor, and somatosensory memories that are used for production. These exemplar banks could pose the source of the variations in Plag et al. (2017). The semantics of *dog + pl* maps to an exemplar *dogs* already in the memory, but the phonological form corresponding to *hoatzin + pl* must be composed from the stem and the plural suffix for people who are not ornithologists. In this novel combination, the undifferentiated /s/ is used.

The situation here bears resemblance to Becker et al. (2011), in which the authors take Turkish voicing alternations as an example of a pattern in the lexicon that is "invisible" to phonological learning. They found that the vowel quality was a strong predictor of voicing alternation in a stem; however, no Turkish speaker used vowel quality as a predictor for when they alternated in nonce words. In essence, the tendency in the spoken lexicon did not generalize to the nonce words, which is exactly the situation here. The authors took their findings as evidence for a universal grammar that is blind to some patterns in the lexicon, but not to others. A similar learning filter could be at work over English S suffixes, where the grammar does not see the variation in length as a possible generalization.

It is also worth noting that most of these S affixes are instantiations of lexical items with many possible phonemic forms. For example, the is-clitic is merely a form of the English copula, which has many realizations, and any theory that ties phonetic realization to semantic identity must account for phonetic instructions of the other realizations. Exemplar theory does this well.

Lastly, a note on the statistical findings of the present study. All of the variations reported in section 10.3 are on the order of individual milliseconds. Though these are statistically significant differences, it is unclear whether they are cognitively significant differences. Word-final /s/ in English has an average duration of 95 milliseconds, meaning the durational changes are miniscule—most English speakers do not notice the difference between the duration of a word-initial s and a word-final s, which amounts to 34 milliseconds (Umeda, 1977). Moreover, the instruments we use measure may not have high enough resolution to make these small millisecond differences useful, which is why Foulkes et al. (2018) suggests, in the interest of scientific conservatism, rounding durations in increments of 5 milliseconds. By this most conservative measure, none of the effects seen here are significant, and, as investigators using statistics, we find ourselves looking at the faces in the clouds rather than signals in the noise.

References

Audacity Team. 2019. Audacity® software is copyright © 1999-2018 Audacity Team. Web site: https://audacityteam.org/. It is free software distributed under the terms of the GNU General Public License. The name Audacity® is a registered trademark of Dominic Mazzoni. https://audacityteam.org/.

Baayen, R Harald, and Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3:12–28.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.

Bauer, Laurie, Rochelle Lieber, and Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford University Press.

Becker, Michael, Nihan Ketrez, and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in turkish laryngeal alternations. *Language* 87:84–125.

Boersma, Paul, and David Weenink. 2019. Praat, a system for doing phonetics by computer [Computer program]. Version 6.0.49. http://www.praat.org/.

Cho, Taehong. 2001. Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica* 58:129–162.

Chomsky, Noam, and Morris Halle. 1968. *The sound patterns of English*. New York: Harper and Row.

Crosswhite, Katherine. 2009. Praat script resources. UCLA Department of Linguistics. URL `http://phonetics.linguistics.ucla.edu/facilities/acoustic/praat.html`.

Drager, Katie. 2011. Sociophonetic variation and the lemma. *Journal of Phonetics* 39:694–707.

Foulkes, Paul, Gerry Docherty, Stefanie Shattuck Hufnagel, and Vincent Hughes. 2018. Three steps forward for predictability. consideration of methodological robustness, indexical and prosodic factors, and replication in the laboratory. *Linguistics Vanguard* 4.

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

Gahl, Susanne, Yao Yao, and Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66:789–806.

Jackendoff, Ray. 1997. *The architecture of the language faculty*. 28. MIT Press.

Jurafsky, Daniel. 2003. *Probabilistic linguistics*, chapter Probabilistic modeling in psycholinguistics: Linguistic comprehension and production, 39–95. Cambridge, MA: MIT Press.

Kim, Heejin, and Jennifer Cole. 2005. The stress foot as a unit of planned timing: Evidence from shortening in the prosodic phrase. In *Ninth European Conference on Speech Communication and Technology*.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59:1208–1221.

Kuha, Jouni. 2004. Aic and bic: Comparisons of assumptions and performance. *Sociological Methods & Research* 33:188–229. URL `https://doi.org/10.1177/0049124103262065`.

Lennes, Mietta. 2017. SpeCT – Speech corpus toolkit for Praat (v1.0.0). First release on Github .

Lindblom, Björn. 1963. Spectrographic study of of vowel reduction. *The Journal of the Acoustical Society of America* 35:1773–1781.

Nooteboom, Sieb G. 1972. Production and perception of vowel duration: A study of the durational properties of vowels in Dutch. Utrecht: University of Utrecht.

Pitt, Mark A, Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume, and Eric Fosler-Lussier. 2007. Buckeye corpus of conversational speech (2nd release). Columbus, OH Department of Psychology, Ohio State University.

Plag, Ingo, Julia Homann, and Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final S in English 1. *Journal of Linguistics* 53:181–216.

R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Tomaschek, Fabian, Ingo Plag, Mirjam Ernestus, and R Harald Baayen. 2018. Modeling the duration of word-final s in english with naıve discriminative learning. Manuscript, University of Siegen/Tübingen/Nijmegen.

Umeda, Noriko. 1977. Consonant duration in american english. *The Journal of the Acoustical Society of America* 61:846–858.

Walsh, Thomas, and Frank Parker. 1983. The duration of morphemic and non-morphemic /s/ in English. *Journal of Phonetics* 11:201–206.

Winter, Bodo. 2013. Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499.

Winter, Bodo, and Andrew Wedel. 2016. The co-evolution of speech and the lexicon: The interaction of functional pressures, redundancy, and category variation. *Topics in cognitive science* 8:503–513.

Yuang, Jiahong, and Mark Lieberman. 2008. Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics*.

APPENDIX: LINEAR MIXED-EFFECTS MODELING

```
MODEL 1: Constructing a model with one fixed effect, speech rate.

Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ speechRate + (1 | subject) + (1 |
word) +      (1 | sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid
 -2212.4  -2175.8   1112.2  -2224.4     3313

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.6341 -0.6659 -0.0108  0.6198  5.0240

Random effects:
 Groups   Name        Variance Std.Dev.
 word     (Intercept) 0.003725 0.06103
 sentence (Intercept) 0.025868 0.16084
 subject  (Intercept) 0.010983 0.10480
 Residual             0.028541 0.16894
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
             Estimate Std. Error t value
(Intercept)  4.894839   0.063808   76.71
speechRate  -0.075344   0.004594  -16.40

Correlation of Fixed Effects:
           (Intr)
speechRate -0.376
```

Model 2: Adding a fixed effect of log(stem duration) to Model 1.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ speechRate + log(stem) + (1 |
subject) +      (1 | word) + (1 | sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid
 -2483.9  -2441.1   1248.9  -2497.9     3312

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8288 -0.6305 -0.0109  0.5998  4.6379

Random effects:
 Groups   Name        Variance Std.Dev.
 word     (Intercept) 0.011913 0.10915
 sentence (Intercept) 0.023691 0.15392
 subject  (Intercept) 0.007902 0.08889
 Residual             0.026155 0.16173
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
             Estimate Std. Error t value
(Intercept)  2.367890   0.160520   14.75
speechRate  -0.047631   0.004685  -10.17
log(stem)    0.427018   0.024908   17.14

Correlation of Fixed Effects:
           (Intr) spchRt
speechRate -0.451
log(stem)  -0.918  0.345

---
```

The model passes our criteria.

```
anova(model1, model2)

Data: daata
Models:
model1: log(morphemeDuration) ~ speechRate + (1 | subject) + (1 |
word) +
model1:     (1 | sentence)
model2: log(morphemeDuration) ~ speechRate + log(stem) + (1 | subject)
```

```
+
model2:     (1 | word) + (1 | sentence)
       Df     AIC     BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
model1  6 -2212.4 -2175.8 1112.2  -2224.4
model2  7 -2483.9 -2441.1 1248.9  -2497.9 273.44      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

MODEL 3: Adding the following consonant as a fixed effect to model 2.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +      (1 | subject) + (1 | word) + (1 | sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid
 -2519.8  -2470.9   1267.9  -2535.8     3311

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8011 -0.6296 -0.0085  0.5987  4.6323

Random effects:
 Groups   Name        Variance  Std.Dev.
 word     (Intercept) 0.0118279 0.10876
 sentence (Intercept) 0.0007242 0.02691
 subject  (Intercept) 0.0076727 0.08759
 Residual             0.0261581 0.16173
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
                     Estimate Std. Error t value
(Intercept)          2.181463   0.152684   14.29
speechRate          -0.044930   0.004474  -10.04
log(stem)            0.431352   0.024746   17.43
followingConsonantr  0.296065   0.016707   17.72

Correlation of Fixed Effects:
            (Intr) spchRt lg(st)
speechRate  -0.452
log(stem)   -0.954  0.340
fllwngCnsnn -0.005 -0.142 -0.031
```

It passes significance criteria

```
anova(model2, model3)
Data: daata
Models:
model2: log(morphemeDuration) ~ speechRate + log(stem) + (1 | subject)
+
model2:     (1 | word) + (1 | sentence)
```

```
model3: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model3:     (1 | subject) + (1 | word) + (1 | sentence)
       Df     AIC     BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
model2  7 -2483.9 -2441.1 1248.9  -2497.9
model3  8 -2519.8 -2470.9 1267.9  -2535.8 37.883      1  7.513e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

MODEL 4: Adding the following consonant as a fixed effect to model 3.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ speechRate + log(stem) + 
followingConsonant +
    cluster + (1 | subject) + (1 | word) + (1 | sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid 
 -2517.9  -2463.0   1268.0  -2535.9     3310 

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-3.8017 -0.6288 -0.0077  0.5988  4.6316 

Random effects:
 Groups   Name        Variance  Std.Dev.
 word     (Intercept) 0.0116924 0.10813 
 sentence (Intercept) 0.0007239 0.02691 
 subject  (Intercept) 0.0076702 0.08758 
 Residual             0.0261582 0.16173 
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
                     Estimate Std. Error t value
(Intercept)          2.192696   0.154646  14.179
speechRate          -0.044922   0.004474 -10.041
log(stem)            0.431404   0.024747  17.433
followingConsonantr  0.296065   0.016704  17.724
clusterTRUE         -0.023128   0.054365  -0.425

Correlation of Fixed Effects:
            (Intr) spchRt lg(st) fllwnC
speechRate  -0.445                     
log(stem)   -0.940  0.340              
fllwngCnsnn -0.005 -0.142 -0.031       
clusterTRUE -0.160 -0.008 -0.016  0.000
---
```

It is NOT a significant effect, so we leave CLUSTER out of future models.

```
anova(model3, model4)
Data: daata
Models:
```

```
model3: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model3:     (1 | subject) + (1 | word) + (1 | sentence)
model4: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model4:     cluster + (1 | subject) + (1 | word) + (1 | sentence)
       Df     AIC     BIC logLik deviance Chisq Chi Df Pr(>Chisq)
model3  8 -2519.8 -2470.9 1267.9  -2535.8
model4  9 -2517.9 -2463.0 1268.0  -2535.9  0.18      1     0.6714
```

MODEL 5 (the baseline model): Adding the recitation order as a fixed effect to model 3.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
    recitationOrder + (1 | subject) + (1 | word) + (1 | sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid
 -2554.3  -2499.3   1286.1  -2572.3     3310

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.0392 -0.6354 -0.0111  0.6006  4.8194

Random effects:
 Groups   Name        Variance Std.Dev.
 word     (Intercept) 0.011026 0.10501
 sentence (Intercept) 0.001251 0.03536
 subject  (Intercept) 0.007661 0.08752
 Residual             0.025834 0.16073
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
                       Estimate Std. Error t value
(Intercept)            2.455834   0.157925  15.551
speechRate            -0.063282   0.005313 -11.911
log(stem)              0.402088   0.025027  16.066
followingConsonantr    0.305817   0.021366  14.313
recitationOrdersecond -0.040307   0.006549  -6.154

Correlation of Fixed Effects:
            (Intr) spchRt lg(st) fllwnC
speechRate  -0.511
log(stem)   -0.951  0.377
fllwngCnsnn -0.011 -0.132 -0.036
rcttnOrdrsc -0.261  0.523  0.173 -0.069
---
```

It passes our criteria, and this becomes our baseline model.

```
Data: daata
Models:
model3: log(morphemeDuration) ~ speechRate + log(stem) +
```

```
followingConsonant +
model3:     (1 | subject) + (1 | word) + (1 | sentence)
model5: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model5:     recitationOrder + (1 | subject) + (1 | word) + (1 |
sentence)
       Df     AIC     BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
model3  8 -2519.8 -2470.9 1267.9  -2535.8
model5  9 -2554.3 -2499.3 1286.1  -2572.3 36.497      1  1.529e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

MODEL 6: Comparing a model with S suffix type as a fixed effect with the baseline.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
    recitationOrder + type + (1 | subject) + (1 | word) + (1 |
sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid
 -2546.8  -2461.3   1287.4  -2574.8     3305

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.0391 -0.6377 -0.0123  0.6008  4.8266

Random effects:
 Groups   Name        Variance  Std.Dev.
 word     (Intercept) 0.0110942 0.10533
 sentence (Intercept) 0.0009339 0.03056
 subject  (Intercept) 0.0076546 0.08749
 Residual             0.0258371 0.16074
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
                       Estimate Std. Error t value
(Intercept)            2.427661   0.159227  15.247
speechRate            -0.061940   0.005284 -11.723
log(stem)              0.405223   0.025027  16.191
followingConsonantr    0.305181   0.018731  16.293
recitationOrdersecond -0.039441   0.006540  -6.031
type3rdsg             -0.018468   0.032048  -0.576
typehas                0.013865   0.032062   0.432
typepl                -0.014587   0.032255  -0.452
typeplPos              0.022177   0.032319   0.686
typepos                0.018739   0.032140   0.583

Correlation of Fixed Effects:
            (Intr) spchRt lg(st) fllwnC rcttnO typ3rd typehs typepl
typplP
speechRate  -0.513
log(stem)   -0.942  0.371
fllwngCnsnn  0.006 -0.150 -0.040
```

```
rcttnOrdrsc -0.261  0.520  0.169 -0.079
type3rdsg   -0.111  0.049  0.001 -0.007  0.025
typehas     -0.119  0.052  0.010 -0.007  0.027  0.501
typepl      -0.094  0.100 -0.029 -0.014  0.054  0.502  0.501
typeplPos   -0.105  0.074 -0.010 -0.010  0.038  0.499  0.499  0.502
typepos     -0.139  0.091  0.025 -0.012  0.048  0.502  0.502  0.504
0.500
---
It does not pass
anova(model5, model6)
Data: daata
Models:
model5: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model5:     recitationOrder + (1 | subject) + (1 | word) + (1 |
sentence)
model6: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model6:     recitationOrder + type + (1 | subject) + (1 | word) + (1 |
model6:     sentence)
       Df     AIC     BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
model5  9 -2554.3 -2499.3 1286.1  -2572.3
model6 14 -2546.8 -2461.3 1287.4  -2574.8 2.5452      5     0.7697
```

MODEL 7: Comparing a model with suffix (all plural, plural-possessive, and possessives) as a fixed effect to the baseline model.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ speechRate + log(stem) + 
followingConsonant +
    recitationOrder + suffix + (1 | subject) + (1 | word) + (1 | 
sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid 
 -2552.5  -2491.4   1286.2  -2572.5     3309 

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-4.0332 -0.6348 -0.0113  0.6015  4.8213 

Random effects:
 Groups   Name        Variance Std.Dev.
 word     (Intercept) 0.011022 0.10498 
 sentence (Intercept) 0.001212 0.03482 
 subject  (Intercept) 0.007666 0.08756 
 Residual             0.025835 0.16073 
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
                       Estimate Std. Error t value
(Intercept)            2.448970   0.158256  15.475
speechRate            -0.062950   0.005326 -11.819
log(stem)              0.402111   0.025025  16.068
followingConsonantr    0.305653   0.021067  14.509
recitationOrdersecond -0.040087   0.006554  -6.116
suffixTRUE             0.009963   0.020973   0.475

Correlation of Fixed Effects:
            (Intr) spchRt lg(st) fllwnC rcttnO
speechRate  -0.514                            
log(stem)   -0.948  0.375                     
fllwngCnsnn -0.009 -0.134 -0.037              
rcttnOrdrsc -0.263  0.524  0.172 -0.071       
suffixTRUE  -0.070  0.085 -0.012 -0.011  0.046
---
It does not pass.
anova(model5, model7).
```

```
Data: daata
Models:
model5: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model5:     recitationOrder + (1 | subject) + (1 | word) + (1 |
sentence)
model7: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model7:     recitationOrder + suffix + (1 | subject) + (1 | word) + (1
|
model7:     sentence)
       Df     AIC     BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
model5  9 -2554.3 -2499.3 1286.1  -2572.3
model7 10 -2552.5 -2491.4 1286.2  -2572.5 0.2216      1     0.6378
```

MODEL 8: Comparing a model with an interaction between type and log(stem duration) to the baseline model, and the baseline + type model.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ speechRate + (type * log(stem)) +
followingConsonant +
    recitationOrder + (1 | subject) + (1 | word) + (1 | sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid
 -2549.1  -2433.0   1293.5  -2587.1     3300

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.9852 -0.6377 -0.0129  0.5980  4.8446

Random effects:
 Groups   Name        Variance  Std.Dev.
 word     (Intercept) 0.0110715 0.10522
 sentence (Intercept) 0.0009329 0.03054
 subject  (Intercept) 0.0076568 0.08750
 Residual             0.0257411 0.16044
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
                       Estimate Std. Error t value
(Intercept)            2.553646   0.244902  10.427
speechRate            -0.062276   0.005284 -11.785
type3rdsg             -0.027176   0.278679  -0.098
typehas                0.043776   0.293700   0.149
typepl                -0.626639   0.289836  -2.162
typeplPos             -0.441939   0.307564  -1.437
typepos                0.258200   0.285672   0.904
log(stem)              0.382861   0.041683   9.185
followingConsonantr    0.304992   0.018719  16.293
recitationOrdersecond -0.039494   0.006530  -6.048
type3rdsg:log(stem)    0.001640   0.049845   0.033
typehas:log(stem)     -0.005355   0.052613  -0.102
typepl:log(stem)       0.108886   0.051491   2.115
typeplPos:log(stem)    0.083226   0.054934   1.515
typepos:log(stem)     -0.043090   0.051135  -0.843
---
```

It does not pass.

```
anova(model5, model6, model8)
Data: daata
Models:
model5: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model5:     recitationOrder + (1 | subject) + (1 | word) + (1 |
sentence)
model6: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model6:     recitationOrder + type + (1 | subject) + (1 | word) + (1 |
model6:     sentence)
model8: log(morphemeDuration) ~ speechRate + (type * log(stem)) +
followingConsonant +
model8:     recitationOrder + (1 | subject) + (1 | word) + (1 |
sentence)
       Df     AIC     BIC logLik deviance   Chisq Chi Df Pr(>Chisq)
model5  9 -2554.3 -2499.3 1286.1  -2572.3
model6 14 -2546.8 -2461.3 1287.4  -2574.8  2.5452      5    0.76967
model8 19 -2549.1 -2433.0 1293.5  -2587.1 12.2644      5    0.03134 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

MODEL 9: Comparing a model with a suffix-stem duration interaction to the baseline model and the model with suffix.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ speechRate + (suffix * log(stem)) +
followingConsonant +
    recitationOrder + (1 | subject) + (1 | word) + (1 | sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid
 -2552.8  -2485.6   1287.4  -2574.8     3308

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.9873 -0.6305 -0.0086  0.5998  4.8325

Random effects:
 Groups   Name        Variance Std.Dev.
 word     (Intercept) 0.011008 0.10492
 sentence (Intercept) 0.001232 0.03510
 subject  (Intercept) 0.007664 0.08754
 Residual             0.025816 0.16067
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
                       Estimate Std. Error t value
(Intercept)            2.579134   0.179564  14.363
speechRate            -0.063438   0.005334 -11.893
suffixTRUE            -0.246945   0.170016  -1.452
log(stem)              0.379176   0.029169  12.999
followingConsonantr    0.305666   0.021219  14.405
recitationOrdersecond -0.040344   0.006554  -6.155
suffixTRUE:log(stem)   0.046058   0.030247   1.523

Correlation of Fixed Effects:
            (Intr) spchRt sfTRUE lg(st) fllwnC rcttnO
speechRate  -0.477
suffixTRUE  -0.477  0.062
log(stem)   -0.960  0.348  0.509
fllwngCnsnn -0.009 -0.133 -0.001 -0.031
rcttnOrdrsc -0.242  0.524  0.027  0.159 -0.070
sffxTRUE:()  0.473 -0.052 -0.992 -0.514 -0.001 -0.021
---
It does not pass
```

```
anova(model5, model7, model9)
Data: daata
Models:
model5: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model5:     recitationOrder + (1 | subject) + (1 | word) + (1 |
sentence)
model7: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model7:     recitationOrder + suffix + (1 | subject) + (1 | word) + (1
|
model7:     sentence)
model9: log(morphemeDuration) ~ speechRate + (suffix * log(stem)) +
followingConsonant +
model9:     recitationOrder + (1 | subject) + (1 | word) + (1 |
sentence)
       Df     AIC     BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
model5  9 -2554.3 -2499.3 1286.1  -2572.3
model7 10 -2552.5 -2491.4 1286.2  -2572.5 0.2216      1     0.6378
model9 11 -2552.8 -2485.6 1287.4  -2574.8 2.3168      1     0.1280
```

MODEL 10 BASELINE: creating a model with isPl as a fixed effect.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ isPl + log(stem) + speechRate +
followingConsonant +
    recitationOrder + (1 | subject) + (1 | word) + (1 | sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid
 -2552.3  -2491.2   1286.1  -2572.3     3309

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.0391 -0.6354 -0.0111  0.6006  4.8193

Random effects:
 Groups   Name        Variance Std.Dev.
 word     (Intercept) 0.011008 0.10492
 sentence (Intercept) 0.001250 0.03536
 subject  (Intercept) 0.007665 0.08755
 Residual             0.025835 0.16073
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
                       Estimate Std. Error t value
(Intercept)           2.4559503  0.1579592  15.548
isPlTRUE             -0.0001659  0.0225771  -0.007
log(stem)             0.4020785  0.0250472  16.053
speechRate           -0.0632844  0.0053213 -11.893
followingConsonantr   0.3058181  0.0213625  14.316
recitationOrdersecond -0.0403078  0.0065524  -6.152

Correlation of Fixed Effects:
            (Intr) iPTRUE lg(st) spchRt fllwnC
isPlTRUE    -0.022
log(stem)   -0.950 -0.040
speechRate  -0.512  0.056  0.374
fllwngCnsnn -0.011 -0.007 -0.036 -0.132
rcttnOrdrsc -0.262  0.031  0.172  0.523 -0.070
convergence code: 0
```

MODEL 10: comparing a model where isPl interacts with log(stem duration) to the baseline model and the model with only isPl.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(morphemeDuration) ~ (isPl * log(stem)) + speechRate + 
log(stem) +
    followingConsonant + recitationOrder + (1 | subject) + (1 | 
word) + (1 | sentence)
   Data: daata

     AIC      BIC   logLik deviance df.resid 
 -2561.1  -2493.9   1291.5  -2583.1     3308 

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-3.9737 -0.6340 -0.0101  0.5980  4.8418 

Random effects:
 Groups   Name        Variance Std.Dev.
 word     (Intercept) 0.010991 0.10484 
 sentence (Intercept) 0.001278 0.03575 
 subject  (Intercept) 0.007657 0.08751 
 Residual             0.025748 0.16046 
Number of obs: 3319, groups:  word, 16; sentence, 12; subject, 10

Fixed effects:
                       Estimate Std. Error t value
(Intercept)            2.647856   0.168072  15.754
isPlTRUE              -0.605124   0.185275  -3.266
log(stem)              0.368125   0.027039  13.614
speechRate            -0.063868   0.005319 -12.007
followingConsonantr    0.305715   0.021577  14.169
recitationOrdersecond -0.040523   0.006543  -6.193
isPlTRUE:log(stem)     0.108103   0.032857   3.290

Correlation of Fixed Effects:
            (Intr) isPTRUE lg(st) spchRt fllwnC rcttnO
isPlTRUE    -0.345                                    
log(stem)   -0.955  0.373                             
speechRate  -0.490  0.035   0.357                     
fllwngCnsnn -0.012  0.001  -0.032 -0.130              
rcttnOrdrsc -0.248  0.011   0.162  0.524 -0.069       
isPlTRUE:()  0.345 -0.992  -0.380 -0.028 -0.002 -0.007
---
```

It does not pass.

```
anova(model5, model10baseline, model10)
Data: daata
Models:
model5: log(morphemeDuration) ~ speechRate + log(stem) +
followingConsonant +
model5:     recitationOrder + (1 | subject) + (1 | word) + (1 |
sentence)
model10baseline: log(morphemeDuration) ~ isPl + log(stem) + speechRate
+ followingConsonant +
model10baseline:     recitationOrder + (1 | subject) + (1 | word) + (1
| sentence)
model10: log(morphemeDuration) ~ (isPl * log(stem)) + speechRate +
log(stem) +
model10:     followingConsonant + recitationOrder + (1 | subject) + (1
|
model10:     word) + (1 | sentence)
                Df     AIC     BIC logLik deviance   Chisq Chi Df
Pr(>Chisq)
model5           9 -2554.3 -2499.3 1286.1  -2572.3
model10baseline 10 -2552.3 -2491.2 1286.1  -2572.3  0.0001      1
0.994322
model10         11 -2561.1 -2493.9 1291.5  -2583.1 10.8044      1
0.001013 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```