To appear as:

Zuraw, Kie; Lin, Isabelle; Yang, Meng; Peperkamp, Sharon. 2020. Competition between whole-word and decomposed representations of English prefixed words. *Morphology* (2020). https://doi.org/10.1007/s11525-020-09354-6

Abstract

English aspiration is influenced by word structure: in general, a voiceless stop following *s* is unaspirated (*des[t]royed*), but it can be aspirated if a prefix-stem boundary intervenes (*dis[t^h]rusts*) (Baker, Smith & Hawkins 2007). In a production study of 110 words prefixed with *dis- or mis-*, we show that even in prefixed words, there is variation (*dis[k]laimers* ~ *dis[k^h]laimers*), and that aspiration in such words is correlated with word and stem frequency. The more frequent the word, the less likely aspiration, but the more frequent the stem, the more likely aspiration. This contrasting frequency effect is characteristic of the type of competition Hay posits between whole-word lexical access and morphologically decomposed lexical access (Hay 2003): frequent words will tend to be accessed as wholes (and therefore behave as though there is no prefix-stem boundary), but frequent stems will encourage decomposed, prefix + stem access. In order to test whether there is active online competition, as opposed to simply frequency effects that are somehow lexicalized, we also conduct a priming experiment. We find that exposing participants to other prefixed words encourages them to aspirate target words, as compared to when they have been exposed to similar but non-prefixed words. These results provide evidence for active online competition.

keywords: English, aspiration, priming

1 Introduction

This study examines which factors influence whether a (historically) morphologically complex word is pronounced as though it is complex, or as though the prefix and stem were fused. We focus on the English prefixes *dis-* and *mis-*, and variation in how words like *mistake* and *discomfort* are pronounced, specifically in whether the stem-initial consonant is aspirated (*dis[kʰ]omfort*) or unaspirated (*mis[t]ake*).

This distinction in pronunciation has been noted before. Raffelsiefen (1999) proposes that words historically prefixed with *dis-* and *mis-* can have two different prosodic structures. If the prefix and stem form separate phonological words, as in *(dis)*$_{PWord}$*(color)*$_{PWord}$ or *(dis)*$_{PWord}$*(comfort)*$_{PWord}$, then the prefix bears secondary stress, because it forms a phonological word, and the stem-initial consonant is aspirated, because it is phonological-word-initial: *dìs[kʰ]ólor*, *dìs[kʰ]ómfort*. If they form a single phonological word, as in *(discover)*$_{Pword}$ or *(mistake)*$_{Pword}$, then the historical prefix is unstressed and the stem-initial consonant is unaspirated: *dis[k]óver*, *mis[t]áke*—in terms of their phonology, such words act as though *mis/dis* and what follows form a single morpheme. An acoustic study reported in Baker, Smith and Hawkins 2007 and Smith, Baker & Hawkins 2012 supports a phonetic difference along Raffelsiefen's lines. The Baker/Smith/Hawkins experiment used scripted dialogues to elicit five Southern British English speakers' pronunciations of eight "transparently prefixed" words (e.g., *mistiming*, *distrusts*) and eight "pseudoprefixed" words, which include both historically prefixed words (e.g. *mistakes*, *discovered*) and non-prefixed words that begin with the phonemic sequence /mɪs/ or /dɪs/ (e.g., *destroyed*, *mysterious*). Among other acoustic differences that were significant, their experiment found that in transparently prefixed words, the prefix, especially its vowel, is longer, supporting the claim of secondary stress; and the voice onset time (VOT) of the following stop is longer, supporting the claim of aspiration. In the studies presented in this paper, we chose to focus on VOT as an indicator of underlying morphological structure.

We will show below that many of these words have variable pronunciation (at least in American English, the dialect spoken by our participants). For example, *displacement* may have an aspirated or unaspirated /p/. We will assume that the aspirated pronunciation occurs when a speaker treats the word as decomposed (*dis* + *[p^h]lacement*) and the unaspirated pronunciation occurs when a speaker treats the word as a single unit (*dis[p]lacement*). There are two possible mechanisms for how this structural difference produces the pronunciation difference; nothing in our study hinges on the choice of mechanism. The first is Raffelsiefen's prosodic analysis: the prefix-stem boundary produces a prosodic-word boundary, making the stem-initial consonant prosodic-word-initial, and in English, prosodic-word-initial voiceless stops are aspirated. By contrast, when there is no morpheme boundary, the /s/ of the prefix joins the onset of the following syllable; in English, only syllable-initial consonants can be aspirated, so the stem-initial consonant must be unaspirated. The second possible mechanism is that when there is a morpheme boundary, the stem is subject to output-output correspondence (Benua 1997 and many others), causing it to be pronounced as it is in isolation, with an aspirated initial consonant.

What determines whether a potentially complex word is treated as a single unit? One factor is frequency. To a first approximation, high-frequency words tend to be less compositional, in terms of both meaning and pronunciation (e.g.(Bybee 1988, 2006)). To take an example from Pagliuca (1976), as cited by Bybee (1985), English words with the prefix *pre-* fall on a continuum. The words that behave the most as truly prefixed are those like *predecease*, where the prefix has its citation vowel [iː], the meaning tends to be straightforwardly composed from the meanings of the parts (*predecease* = 'to die before'), and frequency tends to be low. The words that behave the most as though the prefix and stem are fused are those like *preface*, where the prefix vowel is [ɛ], the meaning tends to be opaque (*preface* has little to do with *face*), and the frequency tends to be higher.

Hay (2003), however, found that the *relative* frequency of the whole word and its component parts better predicts compositionality. She presented participants with complex word pairs that had different stem frequencies, but were matched in surface frequency (as well as other prosodic and phonotactic characteristics). Participants judged which word in each pair was more decomposable. She found that words that are more frequent than their stems, like *refurbish*, which is much more frequent than *furbish*, were rated less complex, while words that are less frequent than their stems, like *rekindle*, were rated more complex. Hay argues that these results are congruent with dual-route models of morpho-phonological processing in which a decomposed access route races against with a direct access route (Wurm 1997; Caramazza, Laudanna & Romani 1988). Higher word frequency would make the direct, whole-word route faster, and higher stem frequency would make the decomposed route faster.

Previous studies have investigated competition between whole and decomposed structure as an explanation for variation in pronunciation (e.g., Baroni 2001; Zuraw 2009; Ben Hedia & Plag 2017—we will return to these in section 4). But does the competition occur in real time during lexical access, or is it fossilized in stored forms? If each *dis*- or *mis*- word is lexicalized as prefixed or non-prefixed, each speaker should pronounce each word consistently, and variation should be possible only between speakers. We could allow stored forms to be variable, for instance with a probability attached to each of the two representations; this would allow within-speaker variation that would be stable within a speaker, not be influenced by external factors such as priming. If the competition instead happens online, as in the dual-route model, we would expect within-speaker variation, and priming should be possible: experimental manipulations that give an advantage to one access route or the other should influence pronunciation.

In the present study, we examine how stem and word frequency influence pronunciation of *dis*- and *mis*- prefixed words in American English, and which mechanism underlies variable pronunciation in individual speakers. We carry out two production experiments. In the first, we use a between-speaker design with a large sample of prefixed and pseudo-prefixed words with *dis*- and *mis*- and show that for many words, speakers vary in whether they aspirate the stop following the prefix. We examine which factors influence whether there is aspiration, a phonetic correlate of morphological structure. The items that show the largest amount of variation (i.e., for which the numbers of aspirated and unaspirated tokens are most similar) are then used in the second experiment. In this experiment, we use a priming paradigm with a within-speaker design to test whether for these varying words we can observe phonetic consequences of competition between whole-word and decomposed representations within individual speakers.

2 Experiment 1: Establishing variation in a large set of words

Our first experiment had two goals. First, we verified that there really are words prefixed with *dis*- and *mis*- whose pronunciation varies between aspirated and unaspirated, and determined which words showed the most variation (for use in Experiment 2). Second, we tested the hypothesis discussed above that *dis*- and *mis*- prefixed words will show less aspiration the more frequent the whole word is, and more aspiration the more frequent the stem is.[1]

We conducted a simple production experiment in which participants read phrases like *his discomfort* and *I mispronounce* from a computer screen, and were audio-recorded. In order not to attract participants' attention to the prefixes *dis*- and *mis*-, the target phrases were interspersed with similar filler phrases, in which the content word started with a different prefix, such as *we reappear*. Stimuli ranged from transparently prefixed words like *mispronounce* to historically prefixed words like *mistake* to completely unprefixed words like *mistress*.

This experiment was reported, in condensed form, in REDACTED.

2.1 Methods

2.1.1 Participants

Participants were 16 members of a university community in Los Angeles, California, nearly all undergraduate students, and all native speakers of American English (5 male, 11 female). Participants completed a survey after the experimental task that asked age, sex, native language(s), and speech- or hearing-impairment status. It then asked participants to state at what age they learned English and to rate their speaking, listening, reading, and writing proficiencies as native, advanced, intermediate, or beginning; finally, participants were asked to fill out the same information for any other languages they spoke. The sixteen participants

[1] The Baker/Smith/Hawkins experiment, with only sixteen words (from fourteen lemmas) was not designed to investigate frequency effects systematically, but the researchers note that the pseudoprefixed words were more frequent and less semantically decomposable, and conjecture that they are treated as single units.

included all reported beginning to learn English by age two (and all but one reported beginning to learn English by age one) and native English proficiency in all four areas.[2]

Four participants reported an additional native language, which we defined as any language learned by age two, even though in half the cases the participant did not report native proficiency in it: Hokkien, Vietnamese (2), and Tagalog. Participants ranged from 19 to 46 years old (median: 21; standard deviation: 7.8). They reported no known speech or hearing impairments, and were paid for their time.

2.1.2 Materials

To select the target *dis/mis* words, we began by extracting from CELEX (Baayen, Piepenbrock & van Rijn 1993) all lemmas beginning with *dis-* or *mis-* that were at least two syllables long. Rather than assembling a small list of words balanced for factors of interest, our goal was to be comprehensive, and include all words of this type, in hopes of getting a good sense of which words can vary. This gave us a full range of words, from those that are clearly not prefixed, not even historically, such as *mister*, to those whose meaning is a transparent combination of prefix and stem, such as *mispronounce*—with many words falling somewhere in between, such as *disconcert* (its meaning, 'unsettle, fluster', has the negative meaning of *dis-*, but bears no relation to the meaning of the much rarer verb *concert*, meaning 'collude').

We strove to include only words that participants would be familiar with, which is challenging since many *dis/mis* words belong to technical or learned registers. To do this, we excluded words with COBUILD lemma frequency of zero, except for a couple that we expected would be familiar, (*discontent* and *miscount*); and excluded a few words that we expected would

[2] One additional participant was excluded for learning English at age 9, and another for equipment failure.

be unfamiliar (despite having COBUILD lemma frequency above zero). For each lemma, such as *dispatch*, we chose one inflectional wordform at random, such as *dispatched*. Though lemma and wordform frequency are highly correlated in these words, we identified several wordforms, like *dispatches* and *mistaken*, that were much more or less frequent than expected from their lemma frequency (as determined by fitting a regression line to wordform frequency as a function of lemma frequency, and looking for outliers). These wordforms were added so that we could test the effects of both lemma and wordform frequency, with the result that a few lemmas were represented by more than one inflectional wordform. We also included *mistook* in order to test our intuition that while other forms of *mistake* would rarely if ever be aspirated, *mistook* would often be aspirated. Finally, to get down to our goal of 110 target words, we excluded some derived wordforms whose semantic relation to a word in our list was transparent. For example, if we already had a word like *dispense*, we might exclude *dispenser*; we stopped excluding such words once we reached the desired number of words. The 110 target words we were left with are listed in Appendix A.

To select 330 filler words (also listed in Appendix A), we chose words that were also prefixed, with *re-*, *in-/im-/il-/ir-*, *con-/com-/col-/cor-*, or *pre-*, such as *react* and *infusion*. Their stems could but did not need to begin with a voiceless stop. To further ensure that the targets didn't stand out from the fillers, we randomly sampled the fillers from CELEX such that they had a similar frequency distribution as the target words, as well as similar distributions of stress pattern (and therefore syllable count), and CELEX morphology code. We removed by hand a few candidate filler words that we judged would be unfamiliar to participants.

Our stimuli included nouns, verbs, adjectives, and adverbs, and we wanted to ensure that participants read each word with the intended part of speech. Stimuli were therefore not just a single word (*mistakes*, ambiguous between noun and verb), but a two-word phrase (*he mistakes*). We also conjectured that reading phrases would produce a more natural prosody than reading just the single word in isolation. The first word in each phrase was a function word from the list *a*, *the*, *some*, *my*, *your*, *his*, *her*, *its*, *our*, *their*, *very*, *to*, *I*, *you*, *he*, *she*, *it*, *we*, *they*, *I'm*, *you're*, *he's*, *she's*, *it's*, *we're*, *they're*, *I've*, *you've*, *we've*, *they've*. For each target or filler word, we identified the set of function words that were syntactically and pragmatically suitable.

Each participant saw a different randomized sequence of 440 stimuli. For each participant, the choice of function word for each target or filler word was randomized, as was the order of the stimuli—with the condition that the first two and last two items always had to be fillers.

2.1.3 *Procedure*

Participants sat alone in a sound-attenuated booth, wearing a head-mounted microphone and looking at a computer screen with a PowerPoint presentation in full-screen mode. Instructions on the screen asked them to read each word or phrase, while speaking naturally and casually and not too carefully, and to press the right arrow key to move from one trial to the next. On each trial, a single phrase, such as *we're disclosing,* was displayed on the screen for the participant to read aloud. The entire session was audio-recorded, using either Audacity (Audacity Team 1999) or PCquirerX (Scicon R&D). It lasted about 20 minutes.

After completing the experiment, the participant signaled to the experimenter, who stopped the recording and gave the participant a paper questionnaire to fill out about age, sex, and language background.

2.1.4 Data exclusion

Out of the 1,760 target stimuli recorded, 44 had to be excluded because the speaker said the wrong target word or the wrong function word, or was disfluent. Many tokens of *discotheque*, in particular, had to be excluded, suggesting that this word is unfamiliar to much of the target population, at least in its written form. Because of experimenter error, one participant was presented with only 109 target items instead of the full 110.

2.1.5 Data coding

For each token, we both measured its voice onset time (VOT) and coded it as aspirated or unaspirated based on expert perception.

To measure VOT, we used Praat (Boersma & Weenink 2006) to manually segment each target word's stem-initial stop closure, and the interval between stop release and onset of voicing in the following sound (VOT interval). We took the end of frication noise of the /s/ in the prefix as the beginning of the stem-initial stop closure. This was indicated by a sharp drop in amplitude in the waveform accompanied by a sharp drop in energy in the spectrogram. The end of the stop closure and beginning of aspiration was taken to be the release of the stop, indicated by the presence of transient noise (a burst). Following Francis, Ciocca and Ching Yu (2003), we took the onset of voicing to be the beginning of periodicity in the waveform. This was done by one author and two research assistants. An example is shown in Figure 1 for the [mɪsk^{h}ælk^{h}] portion of *our miscalculations*; the stem-initial /k/'s closure interval is labeled as *closure*, and its VOT interval is labeled with the whole stimulus phrase for convenience in later processing scripts.

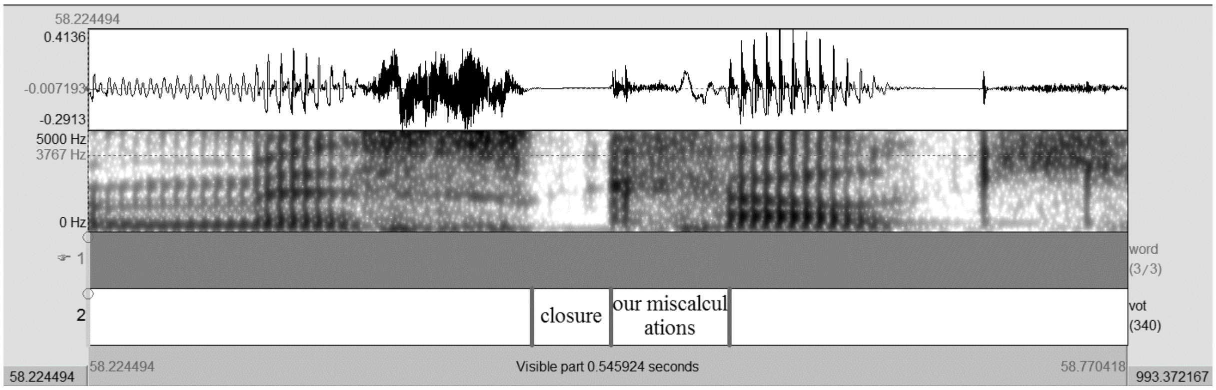

Figure 1: Segmentation of a target item from Experiment 1

Some recorded tokens had no clear release of the stop closure, and therefore VOT could not be measured.

The coding of the stem-initial stop as aspirated or unaspirated was done by a single native-English-speaking author, who performed the task by opening a recording in Praat, and listening to the stem portion of each target stimulus. For example, in *I disprove*, the coder would listen to just *prove*. If the stem sounded like an English word (or non-word) beginning with a voiceless stop, the item was coded as aspirated. If it sounded like an English (non-)word beginning with a voiced stop, the item was coded as unaspirated, since the English voiced stop series is typically pronounced as voiceless unaspirated when word-initial (e.g., Sundara 2005 and references therein). In some cases the item was coded as "unsure". Even if an item was labeled as having no clear closure release, we still attempted to code it perceptually as aspirated or unaspirated. The reason for listening to just the stem was to circumvent the coder's contextual expectations of whether a stop should be aspirated or not based on stress and syllable position. Because aspiration is predictable (non-phonemic) in English, a listener—even one with phonetic training—may impose top-down expectations. Whalen, Best & Irwin (1997) found that English-speaking listeners discriminated [p] and [p^h] poorly, both in a stress context where [p] is expected and in one where [p^h] is expected. By creating a context (word-initial) where both aspirated and unaspirated stops are phonologically legal, we maximized the possibility for veridical perception.

The same author coded each item's stress pattern, since many words vary, such as *discards*, pronounced *discárds* by 14 participants (primary stress on second syllable) and as *díscàrds* by two (primary stress on first syllable, secondary stress on second syllable).

For the analyses of the VOT data below, we excluded any tokens where one or more of the three coders could not identify a stop release (N = 227). For the analyses of the aspirated/unaspirated data, we excluded any tokens coded as "unsure" (N = 171). Stems beginning with /tɹ/ had a strong tendency for the /t/ to be aspirated even where there is no productive prefix boundary and the initial syllable of the "stem" is unstressed – as in *distribution*, for which 10 of 16 tokens were aspirated. Though longer VOT for /tɹ/ is not unexpected for other phonetic reasons,[3] we chose to exclude all stems beginning with this cluster for both types of analyses (VOT and aspirated/unaspirated) to avoid introducing another variable. There were 204 such tokens (120 of which were excluded from the VOT analysis already because of lack of clear stop release). We were left with 1,475 tokens for the VOT analysis below and 1,364 tokens for the aspirated/unaspirated analysis.

---

[3] For example, Zue 1976 found an average VOT of about 25 msec. for English /t/ in utterances like /həˈstɑt/, a non-aspirating environment, and about 35 msec. in utterances like /həˈstɹɑt/, though this was still far less than the VOT for aspirated /t/ in an utterance like /həˈtɑk/, about 70 msec. Similarly, Klatt 1975 found that /t/'s VOT in English words beginning with /stV/ was on average 23 msec, and in words beginning /stɹV/ it was 35 msec (vs. 65 msec for words beginning with /tV/).

2.2 Results[4]

We were interested in how much variation we would find, and in how aspiration would be affected by the frequency of the whole word and the frequency of the word's stem. We also wanted to control for factors such as word length, stress, and stem-initial consonant.

Looking first at the binary aspirated/unaspirated data, overall 42% of tokens were coded as aspirated, 48% as unaspirated, and 10% as "unsure". The histogram in Figure 2 shows how many words were produced aspirated by various numbers of speakers. We can see that most words have either a low rate or a high rate of aspiration—meaning they were aspirated by very few or very many participants, respectively—but many also had intermediate rates.

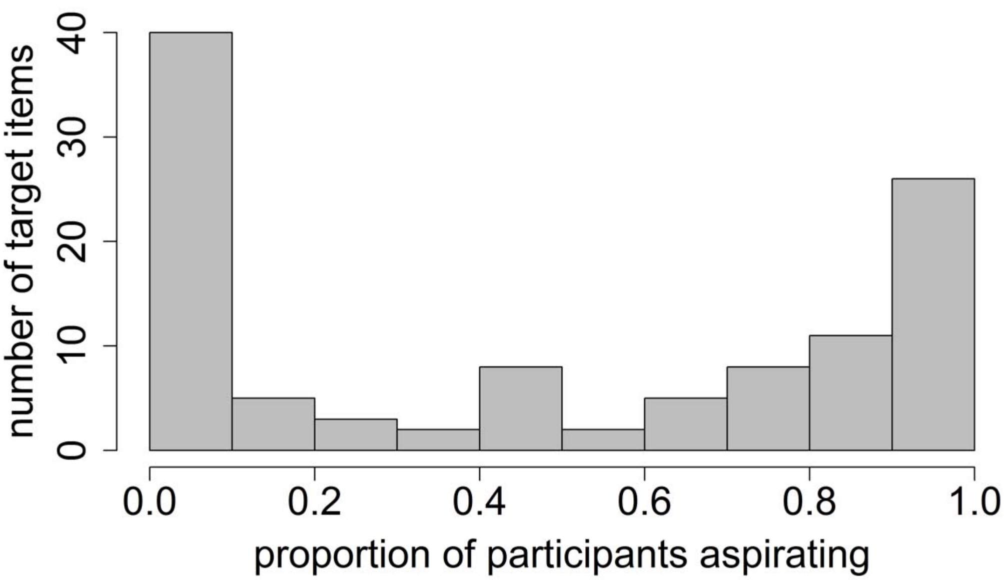

Figure 2: Histogram of aspiration rates

[4] Because we added SUBTLEX-us frequencies to our data and performed model selection anew, the regression models reported here are not the same as those reported in REDACTED, but overall the results are very similar.

Appendix A breaks down the number of aspirated and unaspirated tokens for each word. Impressionistically, we found that, as expected, productively prefixed words had a higher rate of aspiration than the rest. We confirmed our intuition that *mistook* would often be aspirated (14 out of 16 tokens), even though *mistaken* and *mistakes* rarely were (2 and 0 out of 16 tokens); presumably this is because the irregular past tense morphology of *mistook* makes the prefix-stem boundary salient.

Table 1 shows the binary aspirated/unaspirated data split by whether the stem occurs as a freestanding word (e.g., *miscount*) or not (e.g., *disturbing*). For convenience, we use the term 'stem' in the analyses below to refer to whatever comes after *mis* or *dis*. Some of the stems that don't exist as freestanding words are bound morphemes, as in *discriminate* (compare *incriminate*, *recriminate*), others are merely the remainder of the word after *mis* or *dis*, as in *mistress*. As can be seen, aspiration was more likely when the stem is itself a freestanding word: among the tokens that could be coded as either aspirated or unaspirated, 57% of the ones with a stem that exists as a freestanding word were coded as aspirated, compared to only in 9% of the ones with a stem that does not exist as a freestanding word.

	aspirated	unaspirated	unsure
words whose stem exists as a freestanding word (N = 1187)	57%	33%	10%
words whose stem does not exist as a freestanding word (N = 526)	9%	80%	11%

Table 1: Effect of stem's status on aspiration

We fitted a logistic regression model, where the dependent variable was a token's binary classification as aspirated or unaspirated. Independent variables are listed below in

(*1*). Model exploration and comparison using AIC (Akaike 1998) found that frequencies from SUBTLEX-us (Brysbaert & New 2009), a corpus of film subtitles, performed better than COBUILD frequencies from CELEX, and that a word's "CD [contextual diversity] count" in SUBTLEX-us, which is the number of films that the word appears in, performed better than the word's total frequency. Model comparison also found that the stem's lemma frequency did not contribute significantly; as it was highly correlated with the stem's word frequency it was left out. The final model included random intercepts for word and participant,[5] and was fitted in R (R Core Team 2017), using the glmer() function in the lme4 package (Bates et al. 2015), plus the update() function in order to achieve convergence.

(1) Independent variables in model of whether tokens are aspirated or unaspirated

factor variables:

- stem-initial consonant (*p*, *t*, or *k*)
- stem-initial stress (primary, secondary, or none)

binary variables:

- prefix (*dis-* or *mis-*)
- whether the stem exists as a freestanding word
- whether the word or the stem has a higher frequency (in SUBTLEX-us)

continuous variables (centered and converted to z-score):

- the log of the word's CD count (number of SUBTLEX-us films it appears in)
- the log of the word's lemma frequency in SUBTLEX-us
- the log of the stem's frequency in SUBTLEX-us as a freestanding word (0 if it does not exist as a freestanding word)[6]
- the ratio of the word's log frequency to its lemma's log frequency

[5] There were no random slopes included. Exploratory plots showed that slopes were very similar across participants for all variables except the interaction of word and stem frequency, but adding a random slope for that interaction produced non-convergence.

[6] We added 1 to all frequencies before taking the log, to avoid taking the log of 0.

interaction:

- the word's CD count (z-score of log) × the stem's frequency as a freestanding word (z-score of log)

The model results are shown in (2).

(2) Logistic regression model of aspiration: positive coefficients favor aspiration[7]

Fixed effect	Estimate	StD. Error	z value	Pr(>|z|)
(Intercept)	-3.33	1.07	-3.12	0.0018 **
prefix				
= dis-	*reference level*			
= mis-	2.08	0.59	3.54	0.0004 ***
consonant				
= *p*	*reference level*			
= t	-0.07	0.66	-0.10	0.9198
= k	1.80	0.51	3.54	0.0004 ***
stem-initial_stress				
= unstressed	*reference level*			
= secondary-stressed	-0.08	0.71	-0.12	0.9065
= primary-stressed	-1.39	0.58	-2.39	0.0169 *
word's CDcount (z-score of log)	**-0.86**	**0.29**	**-2.96**	**0.0030** **
word's lemma frequency (z-score of log)	**-1.03**	**0.27**	**-3.79**	**0.0001** ***
stem exists as freestanding word?	**2.50**	**1.02**	**2.45**	**0.0144** *
stem's frequency (z-score of log)	**1.04**	**0.47**	**2.20**	**0.0276** *
is word more frequent than stem?	0.84	1.17	0.72	0.4727
word's CDcount × stem's frequency	-0.42	0.26	-1.62	0.1048

[7] We used the vif() function in R's *car* package (Fox & Weisberg 2010) to check the model's degree of multicollinearity. VIF (variance inflation factor) values of 10 or larger are generally considered problematic. The largest VIF value in our model was 2.2, indicating that multicollinearity was not a serious problem for this model.

The significant factors of interest are bolded. The more frequent the word (as measured by either the number of films the word appears in, or the word's lemma frequency), the less probable aspiration is (negative coefficients). If the stem exists as a freestanding word, aspiration is more probable, and the more frequent the stem is as a freestanding word, the more probable aspiration is. (Recall that we defined 'stem' as the remainder of the word after *dis/mis*, even if the word is not prefixed, not even historically (e.g., *disco*, *mistress*)).

Results were similar when we excluded the four participants who reported an additional native language, except that with less data, the effect of the stem's frequency fell below significance ($\beta = 0.81$, $p = 0.12$).

Figure 3 shows the model's predicted probabilities of aspiration[8] as a function of whether the stem exists as a freestanding word and of word frequency, as measured by CDcount (left panel) or lemma frequency (right panel). There is a higher probability of aspiration when the stem exists as a freestanding word, and higher word frequency discourages aspiration.

[8] Using the plot_model() function in R's sjPlot package (Lüdecke 2018), which relies on the ggplot2 package (Wickham 2016). We used the option type="pred", which gives predicted values, with discrete predictors "held constant at their proportions" (sjPlot documentation).

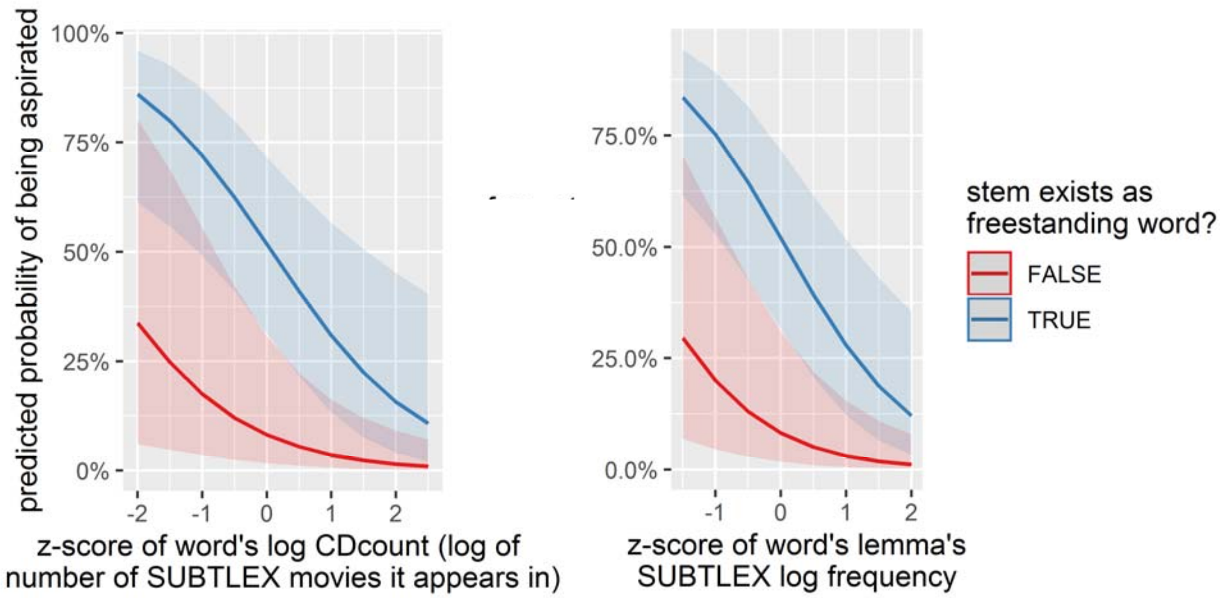

Figure 3: Effect of word frequency on aspiration probability. Shading indicates 95% confidence interval.

Figure 4 shows, for the tokens in which the stem exists as a freestanding word, the model's predicted probabilities as a function of the frequency of the stem as a freestanding word. The greater this frequency, the greater the probability of aspiration.

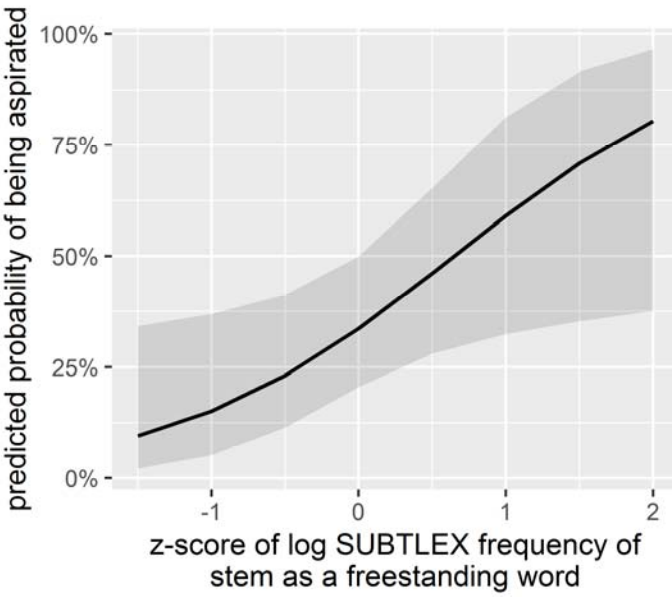

Figure 4: Effect of stem frequency on probability of aspiration

Figure 5 shows the interaction of word frequency (CD count) and stem frequency. Regardless of stem frequency, probability of aspiration is very low when word frequency is high. For lower word frequencies, aspiration probability increases with increasing stem frequency.

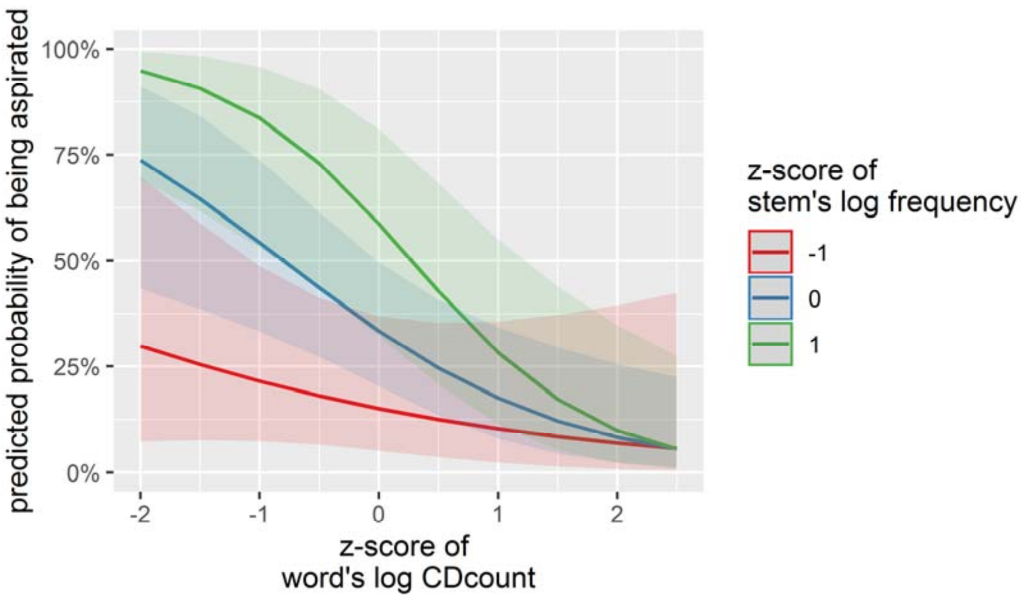

Figure 5: Interaction of word frequency and stem frequency

Turning to voice onset time (VOT), we found that overall it was not bimodally distributed, as can be seen in the left histogram in Figure 6. This is presumably because so many factors go into determining VOT in both aspirated and unaspirated consonants, such as stress, place of articulation, following sound, speech rate, and individual speaker characteristics. Listeners can normalize for these factors, but simply measuring VOT does not result in two clear groups (unaspirated and aspirated). Still, as shown in the right histogram in Figure 6, there is relatively little overlap in VOT between the tokens coded as unaspirated and those coded as aspirated.

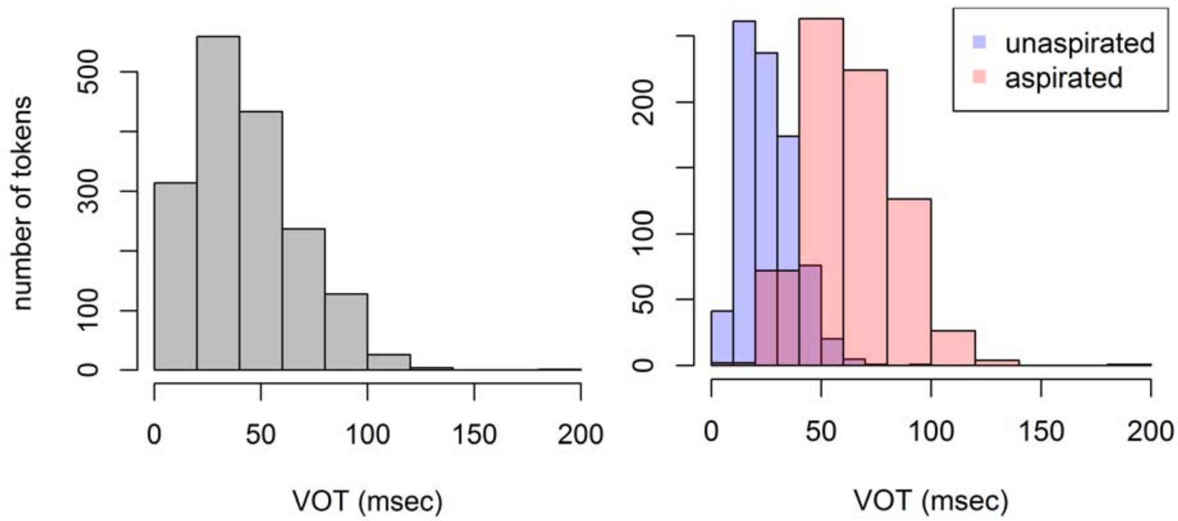

Figure 6: Distribution of voice onset time overall, and within tokens coded as unaspirated or aspirated

The VOT data were modeled with a linear regression. The dependent variable was the logarithm of the VOT, in seconds. We used the log because the raw values were highly skewed. The fixed independent variables are listed in (3). Model exploration and selection, again using AIC, found that unlike in the binary model above, SUBTLEX-us word frequency was a better predictor than SUBTLEX-us's CDcount. Model selection also eliminated some predictors that were included in the binary model: prefix (*dis*- vs. *mis*-), the word's lemma frequency, and whether the word or the stem has a higher frequency. We included participant and word as random effects, as well by-participant random slopes for word frequency and stem frequency.

(3) Independent variables in model of VOT

factor variables:

- stem-initial consonant (*p*, *t*, or *k*)
- stem-initial stress (primary, secondary, or none)
- sound following stem-initial consonant (*j*, *l*, *ɹ*, *w*, or vowel)

binary variable:

- whether the stem exists as a freestanding word

continuous variables (centered and converted to z-score):

- the log of the word's frequency in SUBTLEX-us
- the log of the stem's frequency in SUBTLEX-us as a freestanding word (0 if it does not exist as a freestanding word)
-

interaction:

- the word's frequency (z-score of log) × the stem's frequency as a freestanding word (z-score of log)

The model was fitted using the lmer() function in the lme4 package. The result is summarized in (4). The *p* values listed come from the Anova() function in the car package (Fox & Weisberg 2010); instead of a *p*-value for each level of a factor (such as consonant), this yields a *p*-value for the factor as a whole.

(4) Regression model of VOT[9]

Fixed effect	Estimate	StD. Error	t-value	*p*
(Intercept)	-3.52	0.30	-11.71	< 0.0001 ***
consonant				
				< 0.0001 ***
= p	*reference level*			
= t	0.57	0.09	6.30	
= k	0.61	0.07	8.51	
stem-initial stress				
				0.22
= unstressed	*reference level*			
= secondary-stressed	0.05	0.07	0.66	
= primary-stressed	-0.04	0.07	-0.61	
next sound				
				< 0.0001 ***
=j	*reference level*			
= l	-0.17	0.30	-0.55	
= r	-0.17	0.30	-0.57	
= V	-0.54	0.29	-1.84	
= w	-0.09	0.33	-0.27	
word's frequency (z-score of log)	**-0.11**	**0.03**	**-3.24**	**0.001** **
is stem a freestanding word?	**0.28**	**0.12**	**2.38**	**0.02** *
stem's frequency (z-score of log)	**0.13**	**0.05**	**2.44**	**0.01** *
word's frequency × stem's frequency	**-0.08**	**0.03**	**-2.37**	**0.02** *

The bolded factors are the significant ones of interest: consistent with what we saw in the binary model, higher word frequency reduces VOT (negative coefficient), while the stem's existence as a freestanding word, as well as its frequency, increase VOT.

[9] The highest VIF value found was 1.73, indicating that multicollinearity was not a serious problem in the model. (See footnote 7).

To understand these effects better, we plot the model's predicted values for VOT (the vertical axis is on a log scale, but with labels in raw milliseconds rather than log seconds). Figure 7 illustrates the interaction of word frequency and stem frequency. For infrequent words, the more frequent the stem the longer the VOT. For frequent words, the effect of stem frequency disappears. (The cross-over seen on the right end of the plot is most likely an artefact: there are no words in our stimuli with word-frequency z-score as high as 2 *and* stem-frequency z-score as low as -1.)

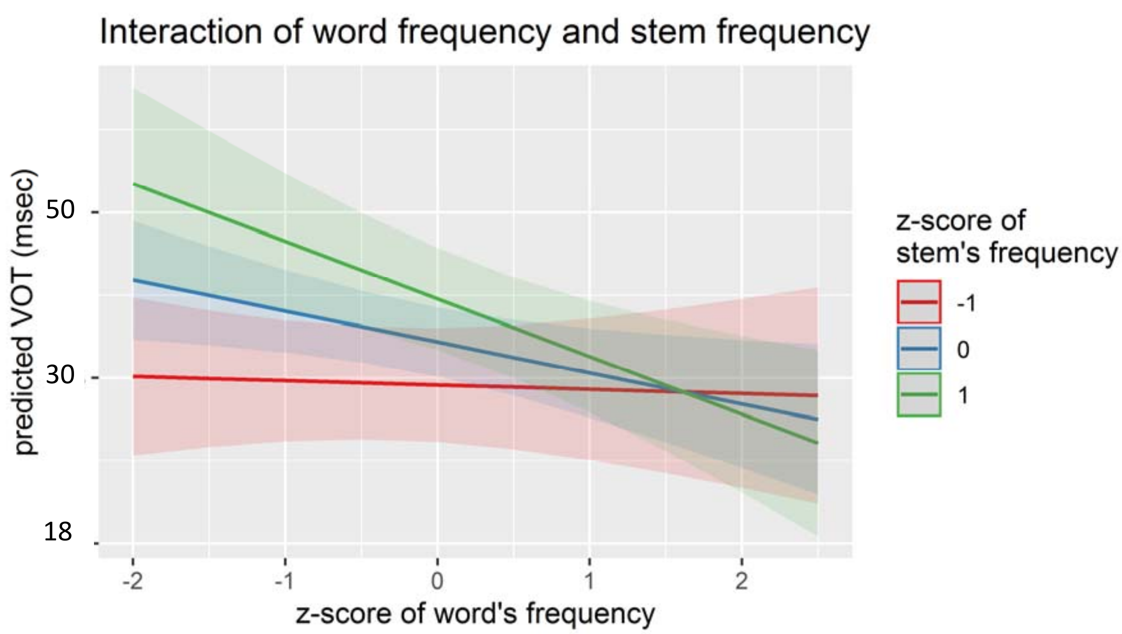

Figure 7: Effect of word frequency and stem frequency interaction on VOT'

When we excluded participants who reported an additional native language besides English, little changed in the model: coefficient values and significance remained similar for each variable.

The results from the binary model (aspirated/unaspirated) and the VOT model are largely congruent. The important overall result is that in both models, higher word frequency suppresses aspiration, and higher stem frequency promotes aspiration.

We explore one discrepancy before moving on: the binary model found that words whose stem-initial syllable is primary-stressed (*discólored*) were less likely to be aspirated than words whose stem-initial syllable is unstressed (*disconnéct*)—and yet, there was no significant effect of stress on VOT. We suspect this is driven by the aspirated tokens: although words like *discólored* were less likely to be aspirated, when they are aspirated (*dis[kʰ]ólored*) their VOTs are longer than an unstressed onset's (*dis[kʰ]onnéct*). To test this conjecture, we fitted separate VOT models to the aspirated tokens and the unaspirated tokens (but without by-participant random slopes, because of convergence issues). We found, supporting our conjecture, a large positive effect of stress on VOT within the aspirated tokens ($\beta = 0.25$ for secondary stress vs. unstressed, $\beta = 0.28$ for primary stress vs. unstressed; $p < 0.0001$), but only a small, non-significant negative effect of stress on VOT within the unaspirated tokens ($\beta = -0.07$ for secondary stress vs. unstressed, $\beta = -0.11$ for primary stress vs. unstressed; $p = 0.17$).

2.2.1 Excluding non-prefixed words

Our rationale for including words that are clearly not prefixed, not even historically, is that there is a continuum from such words (*mistress*), through words that seem prefixed because their stem can combine with other prefixes (*discriminate*, cf. *incriminate*, *recriminate*), through words that lack a transparent semantic relationship to their stem but nonetheless include a negative meaning consistent with the *mis-* or *dis-* prefix (*mistake, disconcert*), to words that are fully transparently prefixed (*discourtesy*). It is not clear how to objectively draw clear distinctions among these categories.

Yet, there is a problem with including words like *mistress* in our statistical analysis: they could be the items driving the effect of whether the stem exists as a freestanding word. We therefore fitted the models above again, excluding these 10 items: *disco*, *discotheque, distance*, *distant*, *distilled, distillery*, *distills, misted, mister*, *mistress*. Our criterion for exclusion was that the etymology in the Oxford English Dictionary's etymology for the word not identify a *mis-* or *dis-* prefix, either in English or in the language the word was borrowed from (French or Latin). (This means we did not exclude words like *dispense*, which were already prefixed when borrowed into English, because they tend to have a semantics that supports a parse of *mis/dis + stem*, or, as with *dispense*, tend to have stems that occur with other prefixes—*expense*, *recompense, propensity*.) By excluding these 10 items, we also exclude all the tokens that were produced with primary stress on the first syllable and no stress in the rest of the word (nearly all tokens of *disco*, *distance*, *distant*, *misted*, *mister*, *mistress*), a pattern that never occurs in the synchronically or historically prefixed words.

After excluding these ten words, the binary model (that predicts probability of aspiration) is similar to the model of the full dataset above, except that the effect of stem frequency falls below significance ($\beta = 0.70$, $p = 0.13$). The model that predicts log VOT differs from the model of the full dataset in that the effect of whether the stem exists as a free-standing word is stronger ($\beta = 0.37$, $p = 0.005$), the effect of stem frequency is weaker and not significant ($\beta = 0.10$, $p = 0.08$), and there is no significant interaction between word frequency and stem frequency ($\beta = -0.05$, $p = 0.19$)

2.3 Discussion

We found that while most words were aspirated by either a low or a high proportion of participants, many also were aspirated by closer to half of the participants. Thus, there was substantial variation across speakers in whether or not aspiration occurred. This is compatible with either a lexicalization account or a dual-route account. Note that these results tell us nothing about whether variation occurs within speakers, and whether variation can be influenced in the lab, a question we will turn to in the next experiment.

As to the effects of frequency on whether or not tokens are aspirated, the dual-route hypothesis predicts that if the whole-word representation and the prefix + stem decomposed representation compete for activation, then the more frequent the word is, the more the whole-word representation should win, and the more frequent the stem is, the more the decomposed representation should win. Based on previous results on the phonetic consequences of the decomposability of *dis-* and *mis-* prefixed words, the whole-word representation should not show aspiration, and the decomposed representation should. Thus, our hypothesis was that higher word frequency would discourage aspiration, and higher stem frequency would encourage aspiration; and this is what we found, in terms of both binary classification (aspirated or unaspirated) and voice onset time.

In models of word recognition, frequency effects are often modeled as differences in resting activation. Higher frequency words have higher resting activation and are therefore more readily accessed, while lower frequency words have lower resting activation and therefore require more effort to access. Under this assumption, the frequency of a word becomes part of its stored representation and plays a role in determining (at least probabilistically) how it is accessed (e.g. whole or decomposed) and how it is produced (e.g. with an unaspirated stop or an aspirated stop respectively).

Under the *automatization* account, sequences of elements (e.g. morphemes, words) are gradually transformed through repeated use to a single chunk that is stored holistically. Thus, whether two or more words in sequence are processed separately or as one unit is determined based on the frequency at which the words co-occur. Evidence for this comes from a study on the reduction of English *don't* occurring in different contexts (Bybee & Scheibman 1999), which found that *don't* was most reduced in contexts where it occurs most frequently, such as after *I* and before *know*. In the case of our items, a prefix + stem combination that occurs frequently should be more likely to be processed as a single unit.

A similar account is that *predictability leads to reduction*, possibly so that less effort is spent on material that a listener can better guess (the Smooth Signal Redundancy Hypothesis, Aylett & Turk 2004): speakers reduce (e.g. fail to aspirate) higher-frequency words since they are usually more predictable from the surrounding context. A word's predictability is its probability of occurring with surrounding words (Jurafsky et al. 2001). Our items had little context to predict from, just a single preceding function word, but indeed the less context there is, the more a word's predictability reduces to its frequency. In addition, Seyfarth (2014) argues that words that are usually predictable (and more-frequency words will tend to be more predictable) get stored with reduced representations, and are therefore reduced even when in an unpredictable context.

Both the automatization account and the reduction-of-the-predictable account can explain why higher-frequency words have less aspiration in our data, and depending on how it is implemented the automatization account could explain why higher-frequency stems have less aspiration. But in both accounts lexical frequency pre-determines the amount of reduction we should expect to see in words. This predicts that a given speaker will produce a given word in a consistent way (always aspirated, always unaspirated). Or, if there is variation in production—the same speaker sometimes aspirates and sometimes does not aspirate the word—then the probabilities of the two variants should be stable within a speaker. In other words, we should not be able to prime speakers to produce one variant or the other. In the dual-route account, on the other hand, experimental conditions that encourage one route or the other should influence pronunciation. This is what we investigate in Experiment 2.

3 Experiment 2: Priming study

Experiment 1 found that many words were pronounced differently across speakers, and that words' pronunciation patterns were correlated with their frequency properties. In our next experiment, we test whether the two pronunciations could result from competition between two modes of lexical access.

Our specific aims were twofold. First, we tested whether there is variation not only between speakers but also within speakers. If there is, that would mean that each person can't have just one stored phonological form for a word that fully determines that word's aspiration. Rather, either aspiration is determined (at least partly) online; or a person can have two stored forms, each with a certain probability of being chosen (Baroni 2001); or a person could have an exemplar cloud from which tokens are drawn at random (e.g., Johnson 1997; Pierrehumbert 2001). Second, if there is within-speaker variation, we tested whether it can be influenced experimentally. Do speakers simply choose a stored phonological form or exemplar at random? Or are speakers' pronunciations affected by the course of lexical access on each occasion, as we have assumed? If lexical access affects pronunciation, it should be possible to influence pronunciation by manipulating lexical access experimentally. We therefore examined whether, for variable words, speakers' choice between aspirated and unaspirated pronunciations can be influenced by priming prefixed or unprefixed word structure and hence encouraging one or the other mode of lexical access. We did this in a production experiment with a block design, using just the most-variable words from Experiment 1: in one block, the target phrases (e.g., *to mispronounce*) were interspersed with phrases containing prefixed words (e.g., *you realign*), while in the other block they were interspersed with phrases containing unprefixed words (e.g., *they pirouette*). Importantly, the stems of the prefixed words did not begin with a voiceless stop; thus, any priming effect would need to happen at an abstract, structural level, and not simply because the prime contained stem-initial aspiration. In order to ensure that participants fully accessed the primes, we added a meaning-judgment task on certain trials.

The participant experience in Experiment 2 was very similar to that in Experiment 1: participants once again saw phrases on the screen and read them aloud (with the additional meaning-judgement task, described below, on some trials). The difference was that they saw each target word with *dis-* or *mis-* twice, once in a block full of other prefixed words (like *unleashed*, *injustice*, *uplifting*), and once in a block otherwise full of unprefixed words (like *matured*, *chaotic*, *pirouette*).[10] We predicted that there would be intra-speaker variation, and, in

[10] In a pilot experiment, we used unproductively prefixed primes (*conjugate*) instead of totally unprefixed primes. Moreover, we did not separate the prime types into two blocks, but instead mixed unproductively and productively prefixed primes within both blocks, to test whether the immediately preceding prime influenced pronunciation. (We also did not include the meaning-judgement task described below.) The results showed that it was indeed possible for the same speaker, in the same experimental session, to pronounce the same word both ways (aspirated and unaspirated). Yet, there was no clear correlation between pronunciation and immediately preceding prime type. In this experiment we therefore strengthened the design: the two prime types are maximally different (productively

particular, that speakers would produce more aspiration on the target items in the block with prefixed primes than in the block with unprefixed primes.

3.1 Methods

3.1.1 Participants

Fifty-two university students in Los Angeles, California, all native speakers of American English, participated for course credit.[11] Since participants in the two experiments were recruited through different channels (paid participants for Experiment 1 and undergraduates from a Psychology Subject Pool for Experiment 2), we cannot guarantee that none of the participants in Experiment 2 had participated in Experiment 1. However, as the two experiments were conducted at least two years apart, even if participants had taken part in both experiments, their participation in Experiment 1 should not have affected their performance in Experiment 2. One participant was excluded because of equipment failure, and five were excluded because of error in the forced aligner. This left 46 usable participants: 32 women and 14 men, ages 18 to 31 (median: 20).

Participants completed a nearly identical post-experiment survey as in Experiment 1. Most participants (33) reported beginning to learn English by age 1; six more by age 2, one by age 3, and two by age 4.[12] Twenty participants reported beginning to learn another language

---

prefixed vs. completely unprefixed) and occur in separate blocks, and a meaning-judgement task was added to ensure participants paid attention to the primes.

[11] An additional three participants were recorded but excluded because their variety of English was non-U.S. (UK or South Asia).

[12] The remaining participant reported beginning to learn English at age 15, which we take to be an error (perhaps 15 months was intended).

by age 3, although only four reported native speaking proficiency in that other language: Spanish (6), Farsi (3), Korean (3), Mandarin Chinese (3), Armenian (2), Cantonese (1), Vietnamese (1), and French (1). In section 3.2.1 below, we compare these twenty early bilinguals to the other twenty-six participants.

3.1.2 Materials

We chose 18 target words that showed the most across-speaker variation in Experiment 1 (as well as the most within-speaker variation in the pilot experiment mentioned in footnote 10). We also added the words *misperception* and *mistaking*. *Misperception* was added to increase the number of stimuli whose stem begins with an unstressed syllable. *Mistaking* was added because in Experiment 1, we observed a high rate of aspiration on *mistook*, in contrast with the low rate of aspiration on *mistaken* and *mistakes* (presumably because the irregular past tense morphology of *mistook* makes the prefix-stem boundary salient); we took Experiment 2 as an opportunity to confirm this with a form of *mistake* that occurs primarily as a verb, *mistaking*, in contrast with *mistakes* and *mistaken*, which can be verbs but more often are a plural noun and an adjective, respectively.

Each target word was matched with two primes, one productively prefixed and one unprefixed (according to our judgment). For example, *misconception* was matched with prefixed *indecision* and unprefixed *saturation*. The range of prefixes was broader than those used in fillers in Experiment 1 — see Appendix B for a full list of stimuli. The triplets of target and two primes were matched for part of speech, stress pattern (and thus syllable count), and rough frequency. With one exception (due to an error), the stems of the prefixed primes did not begin with a voiceless stop. As in Experiment 1, we wanted to avoid drawing participants' attention to the *dis-* and *mis-* words, and therefore included 160 additional items as fillers. Half of these were prefixed (with prefixes other than *dis-* or *mis-*) and the other half were unprefixed. Though they were not matched with specific target items, these filler items still served, as a group, to prime the desired morphological structure. We aimed to have a similar distribution of frequencies and stress patterns in the fillers as in the targets and primes; this had the welcome result that the items all come from the same register.[13]

As in Experiment 1, each word (target, matched prime, or filler) was associated with a set of preceding function words, and actual stimuli were two-word phrases (*our misconception*, *her indecision*, *his saturation*), with the function words randomized for each participant.

3.1.3 Procedure

As in Experiment 1, each participant sat alone in a sound-attenuated room or booth. Stimuli were presented one at a time on a screen. Instead of PowerPoint, stimuli were presented using Pygame (Shinners 2000).

Stimuli were presented in two blocks: prefixed and unprefixed. In the prefixed block, participants saw 20 target items, 20 prefixed primes, and 80 prefixed fillers. These were presented in pseudorandom order such that each target was immediately preceded by its designated prime, and the first and last two items in the block were fillers. In the unprefixed block, participants saw the same 20 target items, but pseudorandomized with 20 unprefixed primes and 80 unprefixed fillers. Half of the participants saw the prefixed block followed by

[13] Median SUBTLEX-us (Brysbaert & New 2009) frequency count of target items was 41 (minimum 2, maximum 138). Median for prefixed primes was 26 (range: 0-362), and for prefixed fillers 5 (0-85). Median for unprefixed primes was 24.5 (range: 3-475), and for unprefixed fillers 14.5 (0-674). T-tests on log frequencies show the following significant frequency differences among sets: targets > {prefixed primes, unprefixed primes}, prefixed fillers < all others.

the unprefixed block, while the other half saw the unprefixed block followed by the prefixed block. Between the two blocks, a slide appeared encouraging the participant to take a short break. The participant read each two-word phrase aloud. Moreover, on all trials with a prime (i.e., all trials that immediately preceded a target trial), as well as on a random selection of approximately 12% of the filler trials, participants were asked to rate the phrase on a five-point scale: definitely negative, somewhat negative, neutral/neither, somewhat positive, definitely positive. Overall, about 27% of trials thus included the rating task (the percentage varied across participants, from 24% to 30%, because the random decision as to whether to include the rating task was made independently on filler trials).

The whole session was recorded using Audacity or PCQuirerX. It lasted about 20 minutes.

3.1.4 Data coding

Audio recordings were forced aligned with the online FAVE forced aligner (Rosenfelder et al. 2011), then hand-checked for errors. (Because of technical difficulties, in a few files research assistants had to manually label each target word, without performing further segmentation inside the word.)

A script extracted the target phrases from each recording, and concatenated them in random order, with no indication of each token's block type or order. Each participant was coded by three coders, who rated each token as aspirated, unaspirated, or unsure. Just as in Experiment 1, this was accomplished in Praat by selecting and listening to the stem only, for example *credits* segmented out of *discredits*; if the initial consonant was aspirated, it should be perceived as a typical English voiceless word-initial stop; if it was unaspirated, it should be perceived as a typical English voiced word-initial stop. Five tokens were accidentally overlooked by one designated coder, so another coder supplied her own judgments (still using the usual block-blinded file). All coders, which included two of the authors, had phonetic training and were native speakers of English (and in some cases an additional language). Coders also marked tokens for exclusion because of disfluency or error (where the participant said a word that was not the displayed stimulus). Each token was coded by two or three different coders. When there were three coders, a token was excluded if at least two of them tagged it for exclusion. When there were two coders and they disagreed on excluding the token, a third coder was added to provide the tie-breaking judgement. There were 92 tokens excluded on the basis of disfluency or speaker error.

A token was marked as aspirated or unaspirated if the majority of coders identified it as such. If there was no majority (i.e., one coder said "aspirated", one said "unaspirated", and one said "unsure") or a majority of coders said "unsure", then the item received no perceptual coding. There were 30 tokens with unbreakable ties, 1398 aspirated tokens, and 320 unaspirated tokens.

We also measured the VOT of the voiceless stop following the prefix *dis*- or *mis*-, using the same method as in Experiment 1. That is, for each token, coders segmented out the interval beginning from the stop release of the target voiceless stop to the beginning of voicing on the following vowel. The duration of this interval was then extracted by script. If a majority of coders agreed that the release could not be identified, the item was not ascribed a VOT.

3.1.5 Data analysis

We performed our analysis on so-called *perfect pairs*: a perfect pair is a pair of tokens of the same word, pronounced by the same speaker, where neither was excluded for disfluency, speaker error, or a tie among coders as to whether the token was aspirated. The 920 recorded and coded token pairs (46 speakers x 20 targets) contained 816 perfect pairs. (In the remaining 104 cases, at least one of a speaker's two tokens of a word was excluded.)

3.2 Results

Participants did engage well in the rating task, as gauged by how well correlated their ratings are. We recorded "definitely negative" as -2, "somewhat negative" as -1, "neutral/neither" as 0, "somewhat positive" as 1, and "definitely positive" as 2. Treating these codes as numbers, we looked at the correlation between each trial's rating (by one participant), and the average of other participants' ratings of the same word. The overall correlation coefficient was 0.76 ($p < 0.0001$); with individual participants' correlation coefficients ranging from 0.52 to 0.91 (median 0.82, all $p < 0.0001$). This reassures us that participants were paying attention to the stimulus words.

Moving on to our hypotheses, we did find within-speaker variation. In 602 pairs out of the 812 perfect pairs, the speaker produced both tokens as aspirated, and in 89 cases the speaker produced both tokens as unaspirated. The remaining 125 cases were what we will call *flip pairs*: the participant produced one token as aspirated and the other as unaspirated, in either order. Nineteen out of 20 stimulus words were flipped for at least one speaker, and 40 out of 46 participants had a flip pair for at least one word. Thus, as predicted, the same speaker could indeed pronounce the same word both aspirated and unaspirated.

Figure 8 shows for each of the 20 target words the number of cases of each pair type (i.e., both aspirated, flip, both unaspirated). Some words have a full 46 perfect pairs; others have fewer. As can be seen, most words had a mix of speakers who pronounced both tokens as aspirated, speakers who pronounced both as unaspirated, and speakers who flipped. The only word on which no speakers flipped was *misconception*. We grouped the words into those that, in our judgment, do not bear a clear semantic relationship to their stem (on the left) and those that do (on the right),[14] and sorted within each group by the number of both-aspirated pairs. Except for the very low rates of aspiration in *mistaking* and *dispatched*, the two groups of words had similar amounts of aspiration. *Mistaking* and *mistook* offer an interesting comparison. Baker, Smith & Hawkins (2007) and Smith, Baker & Hawkins (2012) used *mistakes* as a pseudo-prefixed stimulus, because of the intuition that *mistake* has at best a tenuous relationship to *take*; and in our results too, *mistaking* had few both-aspirated pairs. But its past tense, *mistook*, had mostly both-aspirated pairs. [15] We speculate that the irregular past tense is responsible: In English, the only verbs that can be irregular are monosyllables, or monosyllables with a prefix, plus some verbs whose synchronic status as prefixed is debatable, such as *mistake* or *become*. Thus, the irregularity of *mistook* encourages speakers to think of it as prefixed, and to aspirate the /t/.

[14] *Discards* was hard to categorize, because *to discard* has both a transparent sense (when playing cards, to remove a card from one's hand) and an opaque sense (to throw away).

[15] Experiment 1 produced similar results: as shown in Appendix A, *mistaken* and *mistakes* were aspirated by 12% and 0% of participants, whereas *mistook* was aspirated by 88%.

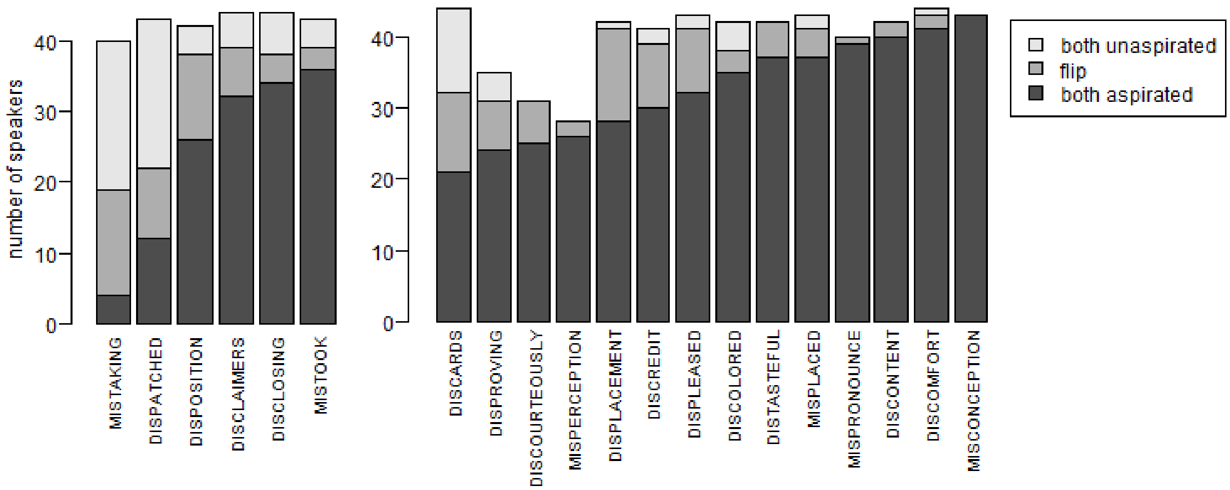

Figure 8: Number of each pair type for each target word. Left: words that do not bear a transparent semantic relationship to their stem; right: words that do bear such a relationship. Some target words have fewer pairs than the maximum of 46, because of discarded tokens.

Figure 11 shows the number of each pair type by speaker, with the speakers sorted by number of aspirated-aspirated pairs. Some speakers had the full 20 perfect pairs, but most had fewer, because of exclusions. We see that most speakers had a mix of words that they pronounced aspirated both times, words that they pronounced unaspirated both times, and words on which they flipped. Only six speakers failed to flip at all. All had at least some aspirated pairs, and 32 out of 36 had at least some unaspirated pairs.

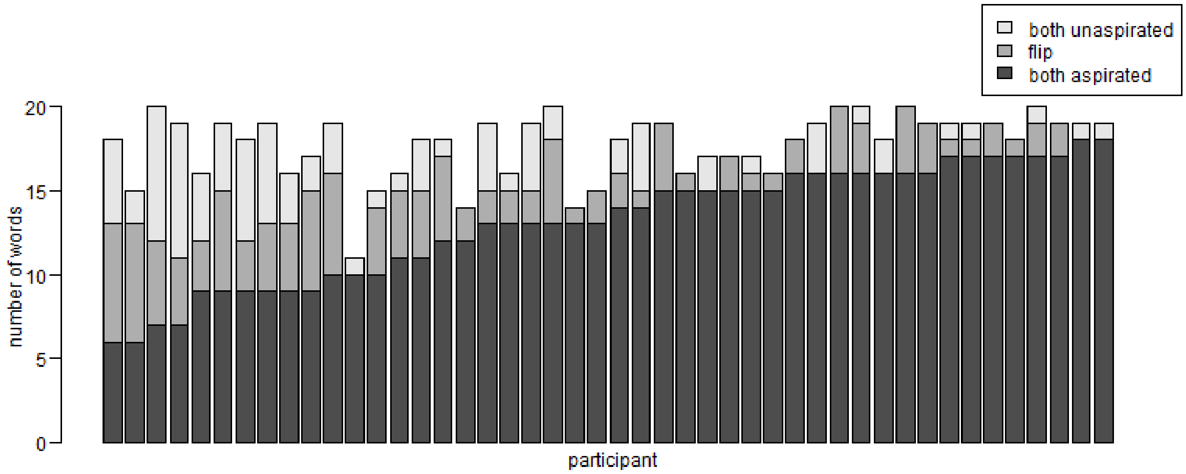

Figure 9: Number of pairs per type, for each participant

The data, then, support our hypothesis that there is within-speaker variation. To test the hypothesis that the variation is influenced by prime type, we must look at whether aspirated tokens were more likely to occur in the block with prefixed primes or the block with unprefixed primes.

Figure 10 shows mean numbers of flip pairs per participant, broken up according to which block the aspirated member of the pair occurred in. In both groups of participants—those who saw the prefixed-prime block first, and those who saw the unprefixed-prime block first—the aspirated token of the flip pair was more often found in the prefixed-prime block (darker bars) than in the unprefixed prime block (lighter bars), consistent with the hypothesis that prefixed primes encourage aspiration.

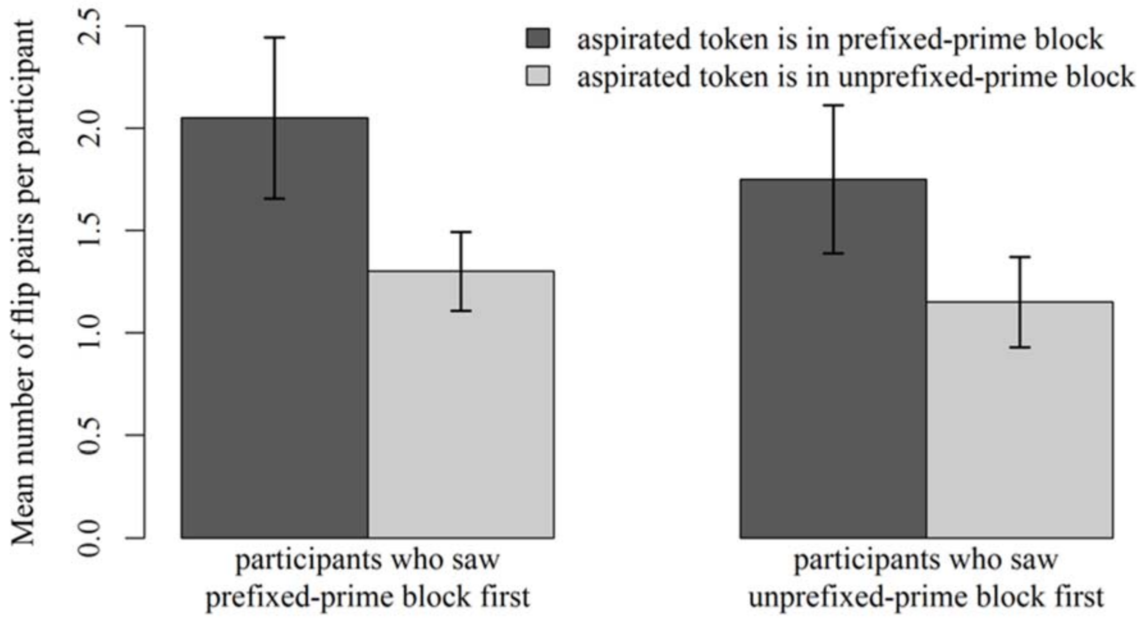

Figure 10: Number of "flip" pairs that occurred in each block

To test the significance of the difference seen in the barplot, we fitted a mixed-effects logistic regression model. The binary dependent variable is whether the aspirated token occurred in the first block (0) or the second block (1).[16] The independent variable is block order (prefixed first or unprefixed first), and we included random intercepts for participant and word. The resulting model (not including random effects) is shown in (5). The null hypothesis would be that block order makes no difference: there might be a baseline preference to aspirate

[16] Alternatively, we could take as the dependent variable whether the aspirated token is in the prefix-prime block versus the unprefixed-prime block, such that the coefficient of interest is the intercept, and add the counterbalancing factor of block order as a fixed effect. The result is the same: no effect of block order, but the intercept is significantly different from zero ($p = 0.04$), indicating that the aspirated member of a perfect pair is significantly more likely to be in the prefixed block, across both block orders.

in the first or second block (intercept), but it shouldn't matter which block had which type of prime. Instead, block order has a significant effect.

(5) Logistic regression model: whether the aspirated token occurred in the first block (0) or the second block (1)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.453	0.318	-1.43	0.154
block order				
prefixed first	*reference level*			
unprefixed first	0.857	0.414	2.07	0.038 *

Because the structure of the model may be unexpected—the independent variable is block order rather than block type (see footnote 16), we step through how the model fits to the data. When the prefixed block comes first, the model predicts a 39% chance that the aspirated token will occur in the second, unprefixed block ($1/(1+e^{0.453})$—that is, the only relevant coefficient is the intercept). This leaves a 61% chance that the aspirated token will occur in the prefixed block, matching the 61% rate in the observed data. When the prefixed block comes second, the model likewise predicts a 60% chance that the aspirated token will occur in the prefixed block ($1/(1+e^{0.453-0.857})$—the intercept and the coefficient for block order both apply), matching the 60% rate in the observed data. In both cases, this is significantly above chance. In summary, in both block orders, the aspirated token is more likely to be found in the prefixed-prime block, as predicted.[17]

[17] With only 20 target words, Experiment 2 was not designed to test frequency effects, and probably lacks the statistical power to do so. At a reviewer's suggestion we added frequency factors to our model (log word CD count, log stem frequency, whether the stem exists as a freestanding word), singly and in a group, and did not find them to have a significant influence. We also tried adding word class to our model (only noun and verb, since there were so few adjectives and adverbs in this model), but it did not

Finally, we also analyzed VOT change within perfect pairs, but the change from the first block to the second block was generally close to zero, with no significant difference between block orders[18]—i.e., it didn't matter if the participant was moving from prefixed to unprefixed blocks, or vice versa. This is presumably because in most perfect pairs, both tokens are aspirated, and thus have similar VOT. The mean absolute value of the VOT difference was 12 msec. in both-aspirated pairs, 6 msec. in both-unaspirated pairs, and 22 msec. in the flip pairs. It would be conceivable for aspirated consonants' VOT to be longer in the prefixed-prime block than in the unprefixed-prime block, but we did not find any such effect.. We also looked at VOT change only within the flip pairs, and there was a non-significant ($p = 0.25$) trend for VOT to increase in participants who start with an unprefixed-prime block and move to a prefixed-prime block, and decrease in participants assigned to the opposite block order. VOT seems to be too noisy to do more than very weakly recapitulate our results for binary coding.

3.2.1 Early bilinguals vs. others

At a reviewer's suggestion, we compared the twenty participants classified as early bilinguals (see section 3.1.1) to the other twenty-six. We found that the two groups were similar in there rates of aspiration and treatment of individual words, but there was a non-significant trend for the early bilinguals to show less of a priming effect.

As Table 2 shows, the rate of flips is similar for early bilinguals and other participants.

---

contribute significantly either. This is not to say that frequency and word class don't matter, but our experimental design makes it unlikely that we could detect their effect.

[18] $p = 0.80$ for block order, using a linear mixed-effects regression model with VOT change from first to second block as the dependent variable

	both tokens aspirated	flip: one token aspirated, one unaspirated	both tokens unaspirated
early bilinguals	258 (73%)	54 (15%)	41 (12%)
others	344 (74%)	71 (15%)	48 (10%)

Table 2: Number of pair types by participant's language background

Of the six participants who never flipped, one was an early bilingual and five were not. Of the fourteen participants who had no unaspirated "perfect pairs", six were early bilinguals and eight were not. (Recall that all participants had at least some aspirated pairs.)

Figure 11 presents the same by-word breakdown as *Figure 8*, but separately for the two participant groups. The two groups' treatment of each word is similar, even including the fact that many tokens had to be discarded for the stimuli *discourteously* and *misperception*.

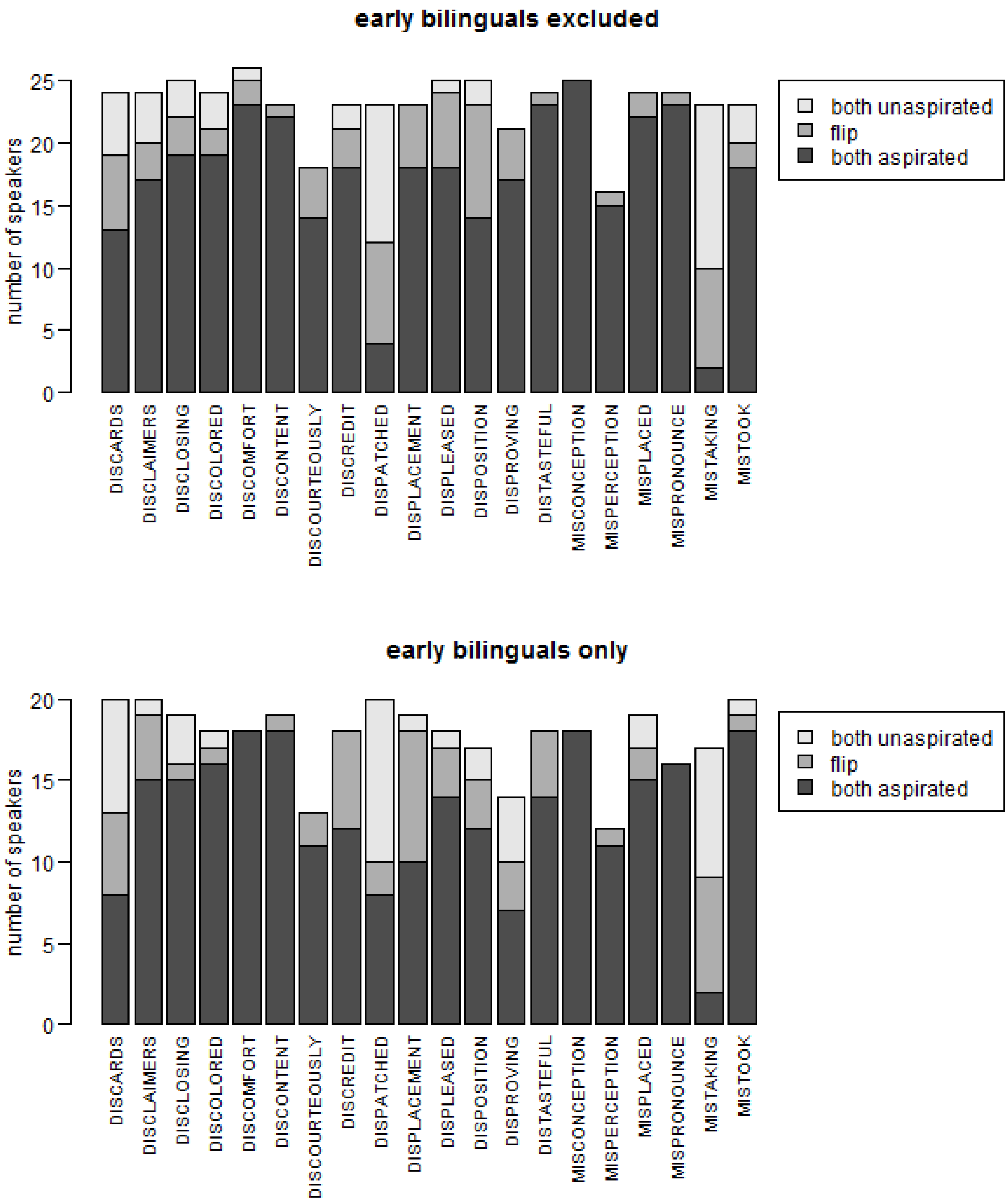

Figure 11: Breakdown of pair types per word, for two types of participant

As Figure 12 shows, there was a trend for the priming effect to be stronger in the non-bilinguals than in the bilinguals, but regression analysis did not find this trend to be significant. Adding early-bilingual status, and its interaction with block order, to the regression analysis in (*5*) found that the effect of block order was still significant ($\beta = 1.28$, $p = 0.03$), there was no significant effect of early-bilingual status ($\beta = 0.30$, $p = 0.62$), and there was no significant interaction between early-bilingual status and block order ($\beta = -0.83$, $p = 0.34$).

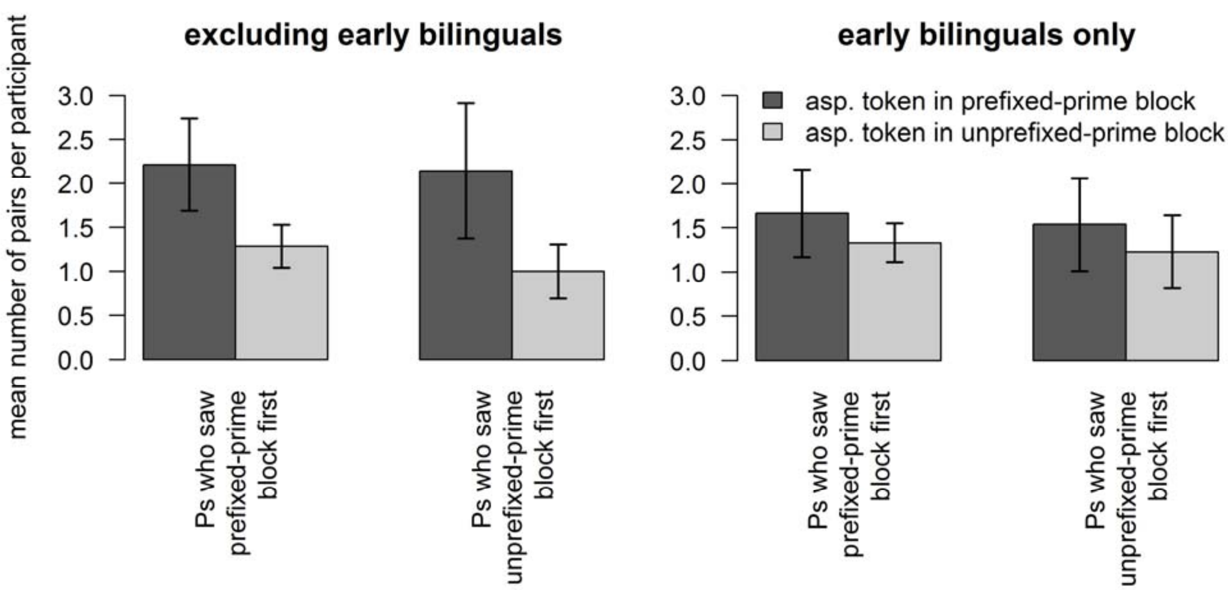

Figure 12: Number of "flip" pairs that occurred in each block, for two types of participant

3.3 Discussion

We found that there is within-speaker variation in the aspiration of *dis-* and *mis-* words, at least for the 20 highly-varying words that we investigated. This held even though we elicited only two tokens of each word from each speaker (so that even if the speaker had maximal, evenly-split variation, we had only a 50% chance of observing the variation), and even though the two tokens were elicited in a single, brief laboratory session. This result is striking given that the phonetic realization of speech is known to be variable but also highly structured. In particular, it is well-established that there is a tight positive linear correlation between the mean VOTs of voiceless stops at different places of articulation in American English for an individual speaker (Chodroff & Wilson 2017 and citations therein). That is to say, speakers are remarkably consistent in their production of VOT even across sound categories when other influencing factors are held constant. However, our results clearly demonstrate that a single speaker can toggle between an aspirated stop and an unaspirated stop even for a single word, lending plausibility to the claim that they are, in fact, accessing different morphological representations online. Thus, we reject the hypothesis that each speaker has a single stored phonological form for each word that fully determines aspiration.

The number of flip pairs, i.e., cases where the same speaker said the same word once aspirated and once unaspirated, was modest (14% of all perfect pairs), but we found that whether the speaker aspirated or not was influenced by the presence vs. absence of words with prefixes other than *dis*- and *mis*-. This goes against random selection of one variant or the other, and supports an online mechanism affecting the variation in aspiration. Because the stems of the primes did not begin with a voiceless stop, priming of the online mechanism must occur at an abstract level—morphological or prosodic structure—and not at the level of allophones ([p~p^h], [t~t^h], [k~k^h]) or phonological features ([+/– spread glottis]).

4 General discussion

In sum, two production experiments were conducted to answer questions about the variability and composition of morphologically complex words in English. In Experiment 1, we aimed to determine whether English *mis*- and *dis*- words could have two pronunciations, and we did find a considerable amount of variation across speakers. By including a large number of target items with different frequency characteristics, and ranging from unprefixed to productively prefixed, we were also able to examine which factors influence whether the words were pronounced as whole (no aspiration) or decomposed (as prefix + stem, with aspiration). Specifically, whole-word pronunciation was more likely as word frequency increased, and the decomposed pronunciation was more likely as stem frequency increased. In Experiment 2, which used just the twenty most-varying words from Experiment 1 (plus two additional ones), we tested whether there is also variation *within* speakers, and if so, whether the choice of whole versus decomposed happens online. We found that, indeed, a word could be pronounced differently by the same speaker, aspirated on one occasion and unaspirated on another; 40 out of 46 participants did this at least once, and it happened at least once for 19 out of 20 words. Further, in support of an online mechanism of access, we were able to influence the likelihood of a word being pronounced as prefixed by priming the speaker with other prefixed words.

Before discussing possible mechanisms for priming, we compare our findings to those in a few other studies that have tied pronunciation differences to a word's decomposability.[19] Ben Hedia and Plag (2017) use corpus data to look at English words with *un*- and *in*- attached to nasal-initial stems. The hypothesis is that if a word is treated as prefixed (e.g., *un+natural*, *im+memorial*), it should have a long [nː] or [mː], but if the prefix and stem are treated as fused (e.g., *immigrant*), the nasal should be short, because long consonants are allowed in English only at morpheme boundaries. They find that words prefixed with *un*- are the most decomposable, according to both semantic measures and relative frequency, followed by

[19] Hanique & Ernestus (2012) argue that studies do not support a role for morphological decomposability in the fine details of phonetic reduction. (More recently, Plag & Ben Hedia (2018) do find effects of decomposability on the durations of *un*- and *dis*-, though not *in*- and *–ly*, in English corpus data.) We focus therefore on studies that, like our own, deal with variation between discrete phonological categories.

negative *in-* (e.g. *incompetent, impossible*), then locative *in-* (e.g. *infuse, import*), and the duration of the nasal reflects this order, being longest in *un-* words and shortest in locative *in-* words. However, they find no effects of decomposability beyond the prefix differences—for example, among words prefixed with *un-*, semantic transparency and relative frequency do not predict the duration of [n]. Thus, unlike in our aspiration data, there is no evidence that individual *un-* or *in-* words' behavior correlates with their frequency properties. Ben Hedia and Plag's corpus method does not permit classifying individual nasal tokens as long or short, so it is not possible to examine variation there. Yet, there is a glimmer of evidence elsewhere for variation in how *un-* and *in-* words are pronounced: Kaye (2005) used a production task to look at a small set of these words, and while most of his ten participants consistently used a long [m] in *immature*, one showed variation.

Baroni (2001) uses a rule in Northern Italian that voices /s/ to [z] between vowels, as in *ca[z]a* 'house', to diagnose word structure. The rule can be blocked at a prefix-stem boundary, as in *pre+[s]elezione* 'preliminary selection', but only if the word is treated as decomposed; fossilized prefixes as in *pre[z]untuoso* 'presumptuous' do not block the rule. Similarly to our aspiration data, Baroni's data shows considerable variation. Many prefixed words showed voicing in at least some tokens, and there are a few words, like *co[s/z]eno* 'cosine', that were highly variable within and across speakers. Baroni finds that the more a word is rated as semantically transparent, the more likely it is that voicing is blocked. Baroni proposes that the lexicon can represent apparently prefixed words as simple or complex, or have two competing representations for the same word.

Zuraw (2009) examines tapping and some other rules in a written corpus of Tagalog, and finds correlations with frequency. Similar to the Northern Italian rule, /d/ usually becomes [ɾ] between vowels, but the rule can be blocked at the prefix-stem boundary, as in *ma+[d]ahon* 'leafy', from *dahon* 'leaf'. In the corpus data, more-frequent prefixed words were more likely to undergo tapping. Zuraw assumes a dual-route model, mediated by prosodic structure. Taken together, these three studies and ours find that words with seemingly similar morphological structure can behave differently depending on whether speakers treat them as truly complex. They also find that the same word can behave differently across speakers and even within the same speaker.

What might be the mechanism for our priming result? That is, why should being exposed to prefixed rather than unprefixed words make participants more likely to give a prefixed pronunciation (with aspiration)? We speculate that the repeated use of a route, direct or decomposed, biases speakers to continue to access subsequent complex words using the same route, potentially overriding the frequency advantage one route normally has over the other for a particular word. During the unprefixed block of the experiment, participants repeatedly were asked to produce unprefixed fillers and primes, necessarily accessing them via a whole-word route (modulo suffixes that might be present). When encountering a target *mis-* or *dis-* word during that block, participants would then be more likely than usual to use the whole-word route again. In the prefixed block, prefixed fillers and primes might be accessed through either route—though probably the decomposed route in most cases—so there would be no bias to use the whole-word route for target items, and probably a bias to use the decomposed route. This bias might come about in several ways. First, it could be the case that (over)use of one route causes speakers to inhibit the use of the other route when accessing morphologically complex words. Alternatively, there could be a cost of switching access modes that increases as one mode continues to be used and the other mode continues in disuse. It is also possible that repeated use of the decomposed route within a short period of time, as would happen in the prefixed block of our experiment, reduces the cost of combining the morphemes that are retrieved, making decomposed access faster than normal.

It is also possible that the priming we observed is not structural or morphological, but rather prosodic. We mentioned in Section 1 that one mechanism to explain why decomposed words have stem-initial aspiration is prosody. Raffelsiefen (1999) and Ogden et al. (2000) both propose that words like *discover* and words like *discolor* have different prosodic structures. For Raffelsiefen, the difference is between one prosodic word, *(discover)*$_{PWord}$ versus two, *(dis)*$_{PWord}$*(color)*$_{PWord}$; the /k/ in *discolor* is then aspirated because it is prosodic-word-initial. For Ogden et al., the difference is that in *discover*, the /sk/ sequence belongs to both the first syllable of the word and the second syllable, whereas in *discolor,* there is a syllable break between the /s/ and the /k/. Taking the Raffelsiefen proposal for concreteness, suppose that the fillers and primes in our unprefixed block all are pronounced with a single-prosodic-word structure, e.g. *(solicit)*, and those in our prefixed block mostly are pronounced with a two-prosodic-word structure, e.g. *(un)(ravel)*. Repeatedly accessing one of these prosodic structures during speech planning could induce participants to favor that structure for the target items in that block. (The other mechanism we mentioned that could explain aspiration in decomposed words, output-output correspondence, would not allow for prosodic priming.)

From our experiments, we cannot determine whether within-speaker variation in aspiration results from a competition between access routes or between prosodic structures. We do need an explanation of why lower word frequency and higher stem frequency favor aspiration, and competition between access routes offers one. But we can't rule out that the frequency factors shape a word's stored form by some other mechanism, and the only online competition is in how to pronounce that form. For example, *discolor*'s frequency profile (low-frequency word, high-frequency stem) would (somehow) result in a stored form with a strong morpheme boundary. The strong boundary would usually result in a two-prosodic-word pronunciation (i.e., aspirated), but priming of single-prosodic-word structure could override that tendency. An account along these lines would require a notion of gradient morpheme-boundary strength in stored forms for complex words; it would also require that the boundary's strength probabilistically determine prosodic structure.

What we can rule out, because we found within-speaker variation, is a model where each word has a single phonological representation. Because of our priming result, we can further rule out a model where a word's representation is probabilistic but fully precompiled—for example, where a speaker has a representation for *discolor* that says it should be treated as complex (and therefore aspirated) 70% of the time, or that it should be divided into two prosodic words (and therefore aspirated) 70% of the time, or even simply that it should be aspirated 70% of the time. Instead, we require an online decision about the word's pronunciation that can be influenced by primes, whether that influence comes in at the level of morphological structure or prosodic structure.

What are some of the factors that could influence aspiration in these *dis- and mis-* words in a more naturalistic setting? Presumably, there might be priming effects at many different levels: priming at the structural level (have there been many prefixed words used recently?), priming at more concrete levels (has an aspirated consonant been used recently?), or priming of a word, prefix, or stem (*color* been used recently?). We speculate that in a corpus study, recent use of a whole prefixed word or of its freestanding stem would be most likely to show effects. In our design, we could test neither of these factors experimentally. As to priming with the stem as a freestanding word, we would face the confound that a prime such as *count* would prime not only the prefixed word *discount* but also the [k^h] pronunciation, due to the presence of the initial aspirated stop. As to priming with the prefix, there are unfortunately too few good primes beginning with *dis-* or *mis-* followed by something other than a voiceless stop.[20] In a corpus study, it might be possible to look for effects of word/prefix/stem priming beyond the effects of phoneme/feature priming. Meaning could also be a factor. For example, when *discard* is used in the context of a card game (taking a playing card from one's hand and returning it to the table), it should be treated as more prefixed than when the word is used more figuratively.

In sum, this study produced several findings that inform our understanding of complex words' pronunciation. First, we showed that both whole-word frequency and stem frequency play a role in determining whether speakers treat morphologically complex words as decomposed or as a whole. We further demonstrated that speakers can be primed to use one structure or the other. These results support a dual-route model of lexical access in which there is active online competition between decomposed and whole-word structure for a morphologically complex word.

5 References

Akaike, Hirotogu. 1998. Information Theory and an Extension of the Maximum Likelihood Principle. In Emanuel Parzen, Kunio Tanabe & Genshiro Kitagawa (eds.), *Selected Papers of Hirotugu Akaike* (Springer Series in Statistics), 199–213. New York, NY: Springer New York. doi:10.1007/978-1-4612-1694-0_15. https://doi.org/10.1007/978-1-4612-1694-0_15 (3 May, 2019).

[20] While there would be many choices of productively prefixed primes, such as *misremember*, there are few choices of unprefixed primes beginning with the same strings for comparison (about twenty-five in CELEX, depending on what one includes), and nearly all of those still suggest the negative meaning of *dis-* or *mis-* even if they are not productively prefixed, such as *disappoint* or *dismantle*.

Audacity Team. 1999. *Audacity*. audacityteam.org.

Aylett, Matthew & Alice Turk. 2004. The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech* 47(1). 31–56. doi:10.1177/00238309040470010201.

Baayen, R. Harald, R Piepenbrock & H van Rijn. 1993. *The CELEX lexical data base on CD-ROM*. Linguistic Data Consortium.

Baker, Rachel, Rachel Smith & Sarah Hawkins. 2007. Phonetic differences between mis- and dis- in English prefixed and pseudo-prefixed words. *Proceedings of ICPhS XVI*, 553–556. Saarbrueken.

Baroni, Marco. 2001. The representation of prefixed forms in the Italian lexicon: Evidence from the distribution of intervocalic [s] and [z] in northern Italian. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1999*, 121–152. Dordrecht: Springer.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Ben Hedia, Sonia & Ingo Plag. 2017. Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics* 62. 34–49. doi:10.1016/j.wocn.2017.02.002.

Benua, Laura. 1997. *Transderivational Identity: Phonological Relations between Words*. University of Massachusetts, Amherst.

Boersma, Paul & David Weenink. 2006. Praat: Doing phonetics by computer, version 4.4. http://www.praat.org/.

Brysbaert, Marc & Boris New. 2009. Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41. 977–990.

Bybee, Joan. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.

Bybee, Joan. 1988. Morphology as lexical organization. In Michael Hammond & Michael Noonan (eds.), *Theoretical morphology: approaches in modern linguistics*, 119–142. San Diego: Academic Press.

Bybee, Joan. 2006. *Frequency of Use and the Organization of Language*. Oxford University Press.

Bybee, Joan & Joanne Scheibman. 1999. The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics* 37(4). 575–596. doi:10.1515/ling.37.4.575.

Caramazza, Alfonso, Alessandro Laudanna & Cristina Romani. 1988. Lexical access and inflectional morphology. *Cognition* 28(3). 297–332. doi:10.1016/0010-0277(88)90017-0.

Chodroff, Eleanor & Colin Wilson. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61. 30–47. doi:10.1016/j.wocn.2017.01.001.

Fox, John & Harvey Sanford Weisberg. 2010. *An R Companion to Applied Regression*. Second Edition. SAGE Publications, Inc.

Francis, Alexander L., Valter Ciocca & Jojo Man Ching Yu. 2003. Accuracy and variability of acoustic measures of voicing onset. *The Journal of the Acoustical Society of America* 113(2). 1025–1032. doi:10.1121/1.1536169.

Hanique, Iris & Mirjam Ernestus. 2012. The role of morphology in acoustic reduction. *Lingue e linguaggio* 11(2). 147–164. doi:10.1418/38783.

Hay, Jennifer. 2003. *Causes and consequences of word structure*. Routledge.

Johnson, Keith. 1997. Speech perception without speaker normalization: an exemplar model. In Keith Johnson & John W Mullenix (eds.), *Talker variability in speech processing*, 145–165. San Diego: Academic Press.

Jurafsky, Dan, Alan Bell, Michelle Gregory & William D Raymond. 2001. Evidence from reduction in lexical production. In Joan L Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 229–254. John Benjamins Publishing.

Kaye, Alan S. 2005. Gemination in English. *English Today* 21(2). 43–55. doi:10.1017/S0266078405002063.

Klatt, Dennis H. 1975. Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters. *Journal of Speech, Language, and Hearing Research* 18(4). 686–706. doi:10.1044/jshr.1804.686.

Lüdecke, Daniel. 2018. *sjPlot - Data visualization for statistics in social science*.

Ogden, Richard, Sarah Hawkins, Jill House, Mark Huckvale, John Local, Paul Carter, Jana Dankovičová & Sebastian Heid. 2000. ProSynth: an integrated prosodic approach to device-independent, natural-sounding speech synthesis. *Computer Speech & Language* 14(3). 177–210. doi:10.1006/csla.2000.0141.

Pagliuca, William. 1976. PRE-fixing. Manuscript. SUNY/Buffalo, ms.

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

Plag, Ingo & Sonia Ben Hedia. 2018. The phonetics of newly derived words: testing the effect of morphological segmentability on affix duration. In Sabine Arndt-Lappe, Angelika Braun, Claude Moulin & Esme Winter-Froemel (eds.), *Expanding the lexicon: linguistic innovation, morphological productivity*, 93–116. De Gruyter.

R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. www.R-project.org.

Raffelsiefen, Renate. 1999. Diagnostics for prosodic words revisited: the case of historically prefixed words in English. In T Alan Hall & Ursula Kleinhenz (eds.), *Studies on the phonological word*, 133-. John Benjamins Publishing Company.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini & Jiahong Yuan. 2011. *FAVE (Forced Alignment and Vowel Extraction)*. fave.ling.upenn.edu.

Scicon R&D. *PCQuirerX*.

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

Shinners, Pete. 2000. *PyGame*. pygame.org.

Smith, Rachel, Rachel Baker & Sarah Hawkins. 2012. Phonetic detail that distinguishes prefixed from pseudo-prefixed words. *Journal of Phonetics* 40(5). 689–705. doi:10.1016/j.wocn.2012.04.002.

Sundara, Megha. 2005. Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *The Journal of the Acoustical Society of America* 118(2). 1026–1037. doi:10.1121/1.1953270.

Whalen, D. H., Catherine T. Best & Julia R. Irwin. 1997. Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics* 25(4). 501–528. doi:10.1006/jpho.1997.0058.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Wurm, Lee H. 1997. Auditory Processing of Prefixed English Words Is Both Continuous and Decompositional. *Journal of Memory and Language* 37(3). 438–461. doi:10.1006/jmla.1997.2524.

Zue, Victor W. 1976. Acoustic characteristics of stop consonants: a controlled study. Technical Report 523. Lincoln Laboratory, Massachusetts Insitute of Technology, ms.

Zuraw, Kie. 2009. Frequency influences on rule application within and across words. *Proceedings of CLS (Chicago Linguistic Society) 43*, 283–309.

Appendix A: Materials for Experiment 1, with results

Target words:

	mean VOT (msec)	standard deviation	number unaspirated	number aspirated	number unsure
discards	39.4	22.1	10	4	2
disclaimers	61.3	26.2	5	10	1
disclosing	50.9	18.1	5	7	4
disclosures	47.9	17.3	6	5	4
disco	24.8	8.3	16	0	0
discolorations	60.2	19.3	2	12	2
discolored	57.7	22.8	5	10	1
discoloring	69.3	25.4	3	12	1
discomfort	51.5	19.2	3	13	0
disconcerting	46.1	15.8	1	12	1
disconcerts	55.4	16.1	0	11	2
disconnect	48.4	13.4	1	13	2
discontent	48.9	18.8	2	14	0
discontinued	45.2	11.8	1	14	1
discontinuity	60.7	20.8	0	13	1
discotheque	29.6	9.1	4	2	2
discounted	62.2	20.6	2	13	0
discourage	41.3	15.0	10	2	4
discourses	54.4	18.0	2	8	4
discourteously	59.8	32.9	7	6	2
discourtesy	62.0	27.9	5	7	3

discovered	25.1	9.2	15	0	1
discredit	60.2	29.3	4	12	0
discredits	63.6	26.6	3	10	2
discreetly	42.9	14.0	12	1	3
discrepancy	28.6	9.0	13	0	2
discretion	30.5	10.6	14	1	1
discriminate	36.2	6.7	10	1	4
discriminated	34.6	10.5	13	0	3
discrimination	34.6	16.2	11	2	3
discuss	27.1	8.3	16	0	0
disparities	15.7	8.9	16	0	0
dispassionate	50.1	24.8	1	14	1
dispatched	22.2	19.6	10	3	3
dispatches	18.0	9.8	12	1	2
dispatching	18.4	12.3	12	2	2
dispel	12.3	5.1	16	0	0
dispelled	14.1	5.3	16	0	0
dispensable	13.9	5.8	14	0	0
dispensaries	13.0	4.5	13	0	2
dispensed	16.5	10.8	15	0	0
dispersed	17.7	5.3	15	0	1
disperses	18.8	12.2	15	1	0
displaced	41.5	23.3	8	5	3
displacement	34.9	18.0	6	6	4
displayed	21.2	10.0	14	0	1
displays	19.1	7.4	15	0	1
displeased	50.8	25.6	8	8	0

displeasing	48.1	26.6	6	6	4
disposable	14.3	5.0	16	0	0
disposal	16.4	5.2	15	0	1
disposed	15.5	7.3	16	0	0
disposition	30.8	13.7	4	7	4
dispossessed	38.2	14.2	1	10	3
disprove	57.9	34.2	5	8	3
disproving	57.2	29.9	3	10	3
disputes	26.7	14.0	13	0	2
disqualified	63.4	21.3	2	12	2
disqualifying	65.0	21.2	2	12	2
distance	29.2	6.8	14	0	2
distant	33.5	13.2	13	3	0
distaste	62.2	20.4	2	13	1
distasteful	60.1	20.7	1	13	2
distastes	55.9	19.8	4	11	1
distemper	49.9	17.1	2	12	1
distilled	29.4	7.8	15	0	1
distillery	32.1	7.2	13	0	3
distills	28.2	7.2	16	0	0
distinctive	23.7	6.1	14	0	2
distinguish	25.8	7.9	14	1	1
distinguished	27.3	8.7	15	0	1
distort	22.5	7.3	13	1	1
distract	34.5	11.7	13	0	3
distressed	47.1	10.3	15	1	0
distressing	35.4	9.4	13	0	3

distributed	56.4	4.4	2	8	5
distribution	45.4	7.2	5	10	1
distributions	46.9	10.4	5	5	5
districts	39.2	8.0	10	3	3
distrust	60.5	28.0	8	7	1
distrustful	58.5	22.3	8	6	2
disturb	29.0	11.8	14	0	2
disturbing	26.7	9.8	15	0	1
miscalculations	65.2	38.0	0	16	0
miscarriages	64.7	18.2	1	15	0
miscarried	56.5	20.9	2	12	2
miscarry	68.9	21.0	0	13	3
misconceived	48.3	15.2	0	13	3
misconceiving	43.5	9.2	1	12	3
misconception	49.1	12.5	2	14	0
misconducted	47.6	11.4	0	16	0
misconducts	52.2	15.8	0	14	1
miscount	67.7	18.6	0	14	1
misplaced	59.3	21.2	1	14	1
misplacing	58.0	18.7	1	12	3
misprint	60.5	18.2	1	15	0
mispronounce	48.6	16.4	1	14	0
mispronunciation	48.3	16.4	1	13	1
misquote	79.5	13.4	0	16	0
misquoted	78.4	19.7	0	15	1
mistaken	26.2	8.5	11	2	3
mistakes	24.5	7.0	15	0	1

misted	29.4	9.3	14	0	1
Mister	31.4	9.2	15	0	1
mistimed	75.7	25.1	0	13	1
mistook	65.4	20.4	2	14	0
mistranslations	80.6	19.3	0	15	1
mistress	39.4	7.8	15	1	0
mistrial	81.2	10.6	0	14	0
mistrusted	55.6	29.2	3	11	2

Filler words:

collapse	considerately	impairing	innovate	reduce	replaceable
collapsed	consignment	impaling	inquired	reductions	replenished
collated	consistencies	imparted	inscription	refinement	replenishments
collating	consolidate	impatient	insecticide	refiner	replica
collectors	conspired	impeached	insomnia	reflector	replicas
collide	constrain	imperatives	inspects	refresher	reported
collided	constraining	imported	inspiring	refrigerators	represses
combined	constricted	imposed	installed	refunded	republicanism
commanders	constrictions	impound	instilling	refute	repugnance
commandment	constructed	impoverished	insulated	refuting	repulsed
commissary	constructive	impregnate	insulted	regained	requesting
commissioned	construing	imprisoned	insure	regalia	required
commodities	consulates	improper	intention	regarded	rerunning
communion	consultant	imprudence	intern	regatta	researching
commuted	consultants	inaccuracies	interned	regattas	resemblance
commuting	consummately	incisors	interning	regress	resembled

compact	contacted	inciting	internment	regressed	reserving
companion	contagiously	include	intimating	regressing	resettlement
compartment	container	incompetency	intruded	regurgitate	resided
compelled	contaminated	incongruous	invade	regurgitates	residual
competitor	contemptible	inconsistencies	invading	rehearsals	resilience
compile	contended	incurs	invariably	reinsuring	resistances
compiles	contestants	indented	inversion	rejoice	resistors
compiling	contextual	indestructibility	invested	rejoicings	resort
compliance	continuation	indifference	investors	relaxes	resorted
complied	continued	indigestibly	invincibly	reliably	resulted
compression	contorted	indigestion	invoke	reliant	resumed
conceded	contorting	individualism	react	relic	resumes
conceitedly	contortions	indoors	reactive	reluctant	resurgence
conceits	contracting	indulged	reactiveness	remand	retainer
conceived	contraction	industrious	reactor	remanded	retention
concentrate	contrive	inert	reacts	remanding	retract
conception	controllable	inertness	reanimate	remands	retractable
concluding	convection	infants	rebutting	remarked	retraction
concocted	conventions	infected	recants	remarking	retrieve
concoction	convergence	infections	receded	remembered	retrieved
condemned	convergent	inference	receiver	reminders	retrieving
condensed	converges	infernos	receptions	remorselessness	reunions
condensing	conversed	infiltrate	recessive	remunerating	reveal
condor	conveyance	infinitives	recital	renaissance	revengeful
confessor	conveyance	inflammatory	recited	renown	reverberating
confided	conveyer	inflection	reciters	rentals	revere
confinements	convulsion	infractions	reclaim	reopen	revered

conflicted	corrections	infringement	recline	reopened	reverent
confrontations	corrode	infringing	reclined	reopens	reversed
congests	corrodes	infuriated	recluse	repaint	reverses
conjectured	corrosiveness	infusion	recruited	repaired	reviewer
conjoined	corrupted	ingenious	recruitment	repairer	reviler
conjunction	corruptible	ingredient	rectories	repairers	revise
conjurers	corruptness	inhabitants	recurred	repayment	revisions
connections	immerses	inhabits	recurrence	repealing	revive
connote	immobilization	inhaler	recycling	repeats	revived
connotes	immobilized	inhibits	redden	repent	revoke
consented	immortalize	iniquity	redeemed	repented	revolve
consider	impaired	injunction	redemptive	repetitive	revolved

Appendix B: Materials for Experiment 2

<u>Target words and associated primes</u>:

target word	*prefixed prime*	*unprefixed prime*
discards	uproots	laments
disclaimers	coequals	placebos
disclosing	enriching	policing
discolored	unfailing	harmonic
discomfort	injustice	momentum
discontent	submarine	volunteer
discourteously	insensitively	sensationally
discredit	unravel	solicit
dispatched	unleashed	matured

displacement	imbalance	charisma
displeased	derailed	patrolled
disposition	malnutrition	punctuation
disproving	enacting	fermenting
distasteful	unhealthy	chaotic
misconception	indecision	saturation
misperception	malformation	liquidation
misplaced	deformed	obeyed
mispronounce	realign	pirouette
mistaking	uplifting	saluting
mistook	withheld	paroled

Prefixed fillers:

befriend	enraged	indelicately	rearmament	unfathomably
begrudging	enrapturing	inelegantly	redeploy	unflinching
coexist	enriched	ineligibly	refashioning	unleashes
coexistence	enriches	ingratitude	reforest	unlikelihood
cohabit	enshrouding	inhabitant	refreshes	unravels
derails	ensnare	insubordination	rehabilitation	unreasonably
dethrone	enthroning	insufferable	reinsure	unseemly
embittered	envenoming	maladjustment	reran	unsightly
embodying	foreshadow	prearrange	subdivide	unswerving
emboldened	foreshadowing	predigest	transforms	untiringly
enable	foreshortening	prerecord	unalterably	unutterable
encircle	forewarn	presuppose	unbeliever	unvarnished
endanger	immoderately	reacts	undoubted	unzip

endangers	immodesty	readjustment	unearthly	unzipped
enfolding	inadequately	reaffirm	unending	upholds
enlarges	indecency	reappear	unfalteringly	withhold

Unprefixed fillers:

ballooning	cements	horizons	monastic	pollutes
battalions	cocoons	hyphenation	negates	questionnaire
benevolently	cultivation	idyllic	neglect	rapacious
bombastic	facetious	imagines	nomadic	rumination
bravado	fallacious	lamenting	nominee	salacious
bureaucracy	fatigued	levitation	oasis	salutes
cadaver	fatigues	libido	obeys	sardonic
calamity	ferments	limitation	pandemic	scholastic
calibration	flamboyant	linguistic	parading	serrated
capacious	flirtatious	magnificent	paroles	significantly
capricious	fluctuation	maneuvers	patrolling	somatic
carbonation	formulation	mature	pedantic	sporadic
caressed	fraternal	melodic	petitioning	thematic
cathartic	frenetic	meticulous	petitions	torrential
cathedral	fumigation	militia	phonetic	vacationing
cavorts	gratuitously	miraculously	polemic	vacations