Interaction of frequency and inflectional status: An approach from discriminative learning

Motoki Saito, Fabian Tomaschek, and R. Harald Baayen

Abstract: High frequency of occurrences has been associated with phonetic reduction on one hand and phonetic enhancement on the other hand. The present study first looks into the possibility that these opposite frequency effects are at least partially due to different inflectional status of the items being investigated. Based on tongue position data from a spontaneous speech corpus of German, we found that stem vowels in inflected words tended to be hyper-articulated (i.e., phonetic enhancement), while those in non-inflected words tended to be articulated with more centralized tongue positions (i.e., phonetic reduction). This observed modulation by inflectional status is subsequently investigated from the perspective of distributional semantics. Using Linear Discriminative Learning to study relations between word embeddings and word forms, we observed that word-final triphones of inflected words received more support from their embeddings (i.e., meanings) compared to non-inflected words. Furthermore, replacement of the two-level factorial predictor of inflectional status with the amount of semantic support led to substantial improvement in model fit. These results suggest direct relationships between word forms and meanings and imply the necessity of such a structure being considered for speech production models.

1 Introduction

Consequences of frequencies of occurrences have been investigated extensively for a wide range of aspects of speech processing, including speech perception and speech production (for an overview, see, e.g., Baayen et al., 2016). And yet it is not entirely clear what frequency and frequency-based measures actually capture.

An influential interpretation of lexical frequency effects in speech production is that higher-frequency words are less informative and that lower degrees of informativity give rise to higher degrees of phonetic reduction. More probable, and less informative, linguistic units such as high frequency words have been found to undergo more phonetic reduction, resulting in shorter acoustic duration (Arnon & Cohen Priva, 2013; Aylett & Turk, 2004, 2006; Bell et al., 2009; Bell et al., 2002; Gahl, 2008; Jurafsky et al., 2001; Pluymaekers et al., 2005a, 2005b), more centralized formant realization (Dinkin, 2008; Wright, 2004), and more reduced tongue positions (Lin et al., 2011; Tomaschek, Arnold, et al., 2018; Tomaschek et al., 2013). According to the smooth signal redundancy hypothesis (Aylett & Turk, 2004), a positive correlation between frequency of occurrences and degrees of phonetic reduction arises due to the cognitive system preferring a stable rate of information in the speech signal. To achieve such a smooth signal, less informative words have to be reduced more.

In contrast, Kuperman et al. (2007) found that more probable interfixes between constituents of Dutch noun-noun compounds were realized with longer duration, rather than shorter duration. They argued that this unexpected positive correlation of probability and phonetic enhancement is paradigmatic in nature. The more probable an interfix is in the paradigm of compounds sharing the same initial constituent, the more the interfix is enhanced in the speech signal. This account is named and referred to as the paradigmatic signal enhancement hypothesis (Cohen, 2014; Tomaschek et al., 2021). The enhancement effects of frequency and paradigmatic probability were subsequently replicated for inflectional suffixes (Cohen, 2014) and for stem vowels of inflected verbs (Tomaschek, Tucker, et al., 2018; Tomaschek et al., 2021).

Why are these opposite directions of frequency effects observed? One possible missing factor is morphological status. The reduction effect of frequency was found when only morphologically simple words were in focus (Lin et al., 2011; Wright, 2004) or when morphologically simple and complex words were not distinguished and therefore aggregated (Aylett & Turk, 2004; Bell et al., 2009; Bell et al., 2002; Dinkin, 2008; Gahl, 2008; Pluymaekers et al., 2005a, 2005b; Tomaschek, Arnold, et al., 2018; Tomaschek et al., 2013). In contrast, the enhancement effect of frequency was found so far exclusively for morphologically complex words (Cohen, 2014; Kuperman et al., 2007; Tomaschek, Tucker, et al., 2018; Tomaschek et al., 2021).

Apart from frequency effects, segments preceding a morphological boundary were found to be acoustically longer (Hay, 2007; Plag & Ben Hedia, 2018; Seyfarth et al., 2017; Smith et al., 2012; Sugahara & Turk, 2009) and articulatorily hyperarticulated (Li et al., 2020; Smith et al., 2012; Song et al., 2013; Strycharczuk & Scobbie, 2016). These findings suggest that phonetic realizations are enhanced before a morphological boundary. Nevertheless, the effect of a morphological boundary and that of frequency have been investigated so far by and large independently. When frequency effects are investigated, morphological status is controlled or simply ignored. When pre-morphological-boundary effects are focused, frequency effects are controlled by item selection (Seyfarth et al., 2017; Sugahara & Turk, 2009), statistically (Plag & Ben Hedia, 2018; Smith et al., 2012) or ignored in some cases (Song et al., 2013; Strycharczuk & Scobbie, 2016). Therefore, it is important to clarify to what extent the enhancement effects of frequency and a morphological boundary are independent from each other and whether the enhancement effect of frequency is confounded with the effect of a morphological boundary. This is the first aim of the current study.

The second aim of the current study is to provide an improved understanding of the pre-morphological-boundary effect. The pre-morphological-boundary effect has mainly been explained in terms of the paradigm uniformity hypothesis (Seyfarth et al., 2017). This hypothesis states that members of the same morphological paradigm are similar to each other in phonetic realization. For example, Seyfarth et al. (2017) found longer duration for stems of inflected words (e.g., *frees*), compared to their corresponding morphologically simple words (e.g., *freeze*).

However, an alternative interpretation of the morphological boundary effect suggests itself within the framework of the discriminative lexicon model (Baayen et al., 2019), a theory that does not require linguistic units such as morphemes, stems, and exponents (Chuang et al., 2020; Stein & Plag, 2021). This approach, which integrates distributional semantics into a computational model for lexical processing, predicts that greater support from word-meanings for their corresponding word-forms goes hand in hand with articulatory strengthening. For example, within this theory, Gahl and Baayen (2024) found that spoken word duration of English homophones was positively correlated with a greater amount of semantic support for word-forms. Well-learned form-meaning relationships are enhanced phonetically, while forms with no support from semantics theoretically predict zero duration (Gahl & Baayen, 2024).

Likewise, Tomaschek et al. (2019) and Tomaschek and Ramscar (2022) have demonstrated similar findings for word final English [s/z] and word final German [ɐ], both of which represent a non-morphemic segment as well as various inflectional functions. Tomaschek and colleagues trained a discriminative learning network to predict the selection of the inflectional functions associated to these segments as well as the selection of the segments themselves. They found that, within these networks, stronger support for the selection was associated with stronger phonetic enhancement of these segments: English [s/z] were longer, and German [ɐ] was pronounced with more pronounced articulatory trajectories as well as more peripheral.

In the light of these findings, we expect that the pre-morphological-boundary effect may in fact reflect different amounts of semantic support that sublexical word-final forms receive from word-meanings. Providing empirical support for this interpretation is the second aim of the current study.

In the following sections, we first address the interaction of morphological status and frequency. Given that previous studies found the enhancement effect of frequency for inflected words (Cohen, 2014; Tomaschek et al., 2021), we also focused on inflected words. We expect that the enhancement effect of frequency persists after including the interaction between frequency and inflectional status in a regression model. Given that a majority of studies reporting the reduction effect of frequency mainly inspected morphologically simple words and that those studies finding enhancement effects of frequency exclusively investigated morphologically complex words, we also expect that phonetic enhancement is present for inflected words, while phonetic reduction is expected for morphologically simple words.

Subsequently, we address the question of a source of the pre-morphological-boundary effect. To this end, we will first introduce a quantitative measure of semantic support based on the discriminative lexicon model, which is expected to be a real-valued alternative for the dichotomy between simple and complex words. We then evaluate this new measure by investigating its predictivity for tongue trajectories registered with electromagnetic articulography. In the discussion section, we discuss possible implications of our results for the understandings of frequency effects in phonetic realizations.

2 Frequency and inflectional status

2.1 Methods

2.1.1 Data

In order to investigate the interaction of frequency by inflectional status, controlling for segmental similarity is essential. It was, however, impossible to find sufficient pairs of morphologically simple and complex words with identical segments over a reasonably wide range of frequencies (e.g., pairs such as *Macht* 'power' vs. *mach+t* 'makes'). Therefore, we extracted all the words with the same rhyme structure with the same nucleus and the same word-final segment, i.e., [a(ː)(X)t], from the articulography section of the Karl-Eberhard-Corpus of spontaneously spoken southern German (KEC: Arnold & Tomaschek, 2016). Our target vowel is [a(ː)], the long and short low open vowels. The word-final segment, which corresponds to a suffix for inflected words, is [t]. To allow an enough number of items to be included, at most one intervening segment was allowed between the target vowel [a(ː)] and the word-final [t]. The resulting set of target words comprised inflected and non-inflected words with and without a morphological boundary between the target vowel and the word-final segment. The stems of the target items comprised not only monomorphemic words but also derived words and compounds. For example, *bemalt* [bəmaːlt] 'paints/painted' consists of a prefix *be-*, a verb stem *-mal-*, and an inflectional suffix *-t*. *Ausland* [aʊslant] 'foreign country' consists of a prefix *Aus-* and a noun *-land*. The former has a morphological boundary between the target vowel [a(ː)] and the word-final [t], while the latter does not. Under this item selection criterion, we were able to collect 560 word tokens from 88 word types, 48 of which were non-inflected and 40 of which were inflected.

For the selected words, vertical tongue tip and body positions were collected from the Karl Eberhards Corpus of spontaneously spoken southern German

(Arnold & Tomaschek, 2016). Since the target vowel is [a(ː)] followed by the word-final [t], the strongest coarticulatory tongue movements were expected for the tongue tip. The tongue body was also included, because a study on coarticulatory tongue movements (Tomaschek, Tucker, et al., 2018) also reported an effect of frequency not only for the tongue tip, but also for the tongue body for words with the stem vowel [a(ː)] and the word-final [t].

Vertical positions of the tongue tip were distributed mainly within -15 mm and +20 mm from the occlusal plane, which was approximated by having the speaker biting a plastic plate (bite plate) (Arnold & Tomaschek, 2016). In some of the word tokens, measurement errors were so big that registered sensor positions jumped around and did not show any consistent pattern of tongue movements. To deal with these jumping data points, intervals between adjacent data points were calculated within each word token. Extraordinarily large intervals which lay outside 1.5 times the interquartile range were considered to be measurement errors, and the data points involved were removed from the data set, where each data point paired time and a vertical tongue tip position. This exclusion procedure amounted to approximately 9.23% of the data points being removed, while the total number of the word tokens was intact. To avoid that the simple removal of the jumping data points leaves too few data points for the time series of tongue positions for a given word token, the word tokens with less than 4 data points after the removal were removed from the dataset. This resulted in the exclusion of 0.39% of the data points and 6.07% of the word tokens.

These word tokens were distributed as shown in Figure 1. The numbers of word tokens spoken by speakers ranged from 1 to 36. The numbers of word types by speakers ranged from 1 to 15. The mean of the numbers of word tokens spoken by each speaker was 15.029, indicated by the vertical line in Figure 1. In Figure 1, different word types are illustrated in different colors and marked by “Word ID”.

For example, speaker “s01” produced the words of interest the most often. Speaker “s35” articulated only one word meeting the criteria of the present study, which is the word with ID “w44”. Word “w44”, *halt*, is listed in the color legend of Figure 1.

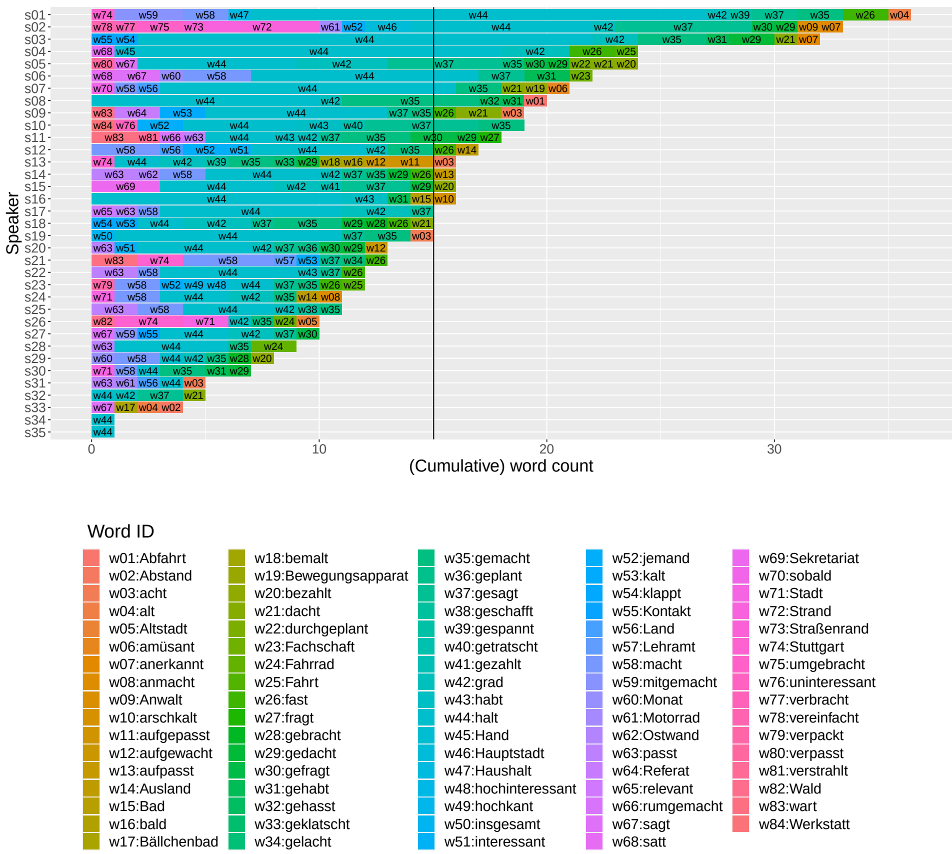


Figure 1: The distribution of the words analyzed in the present study across speakers.

2.1.2 Analysis

The tongue positions during [a(ː)] were fitted with Generalized Additive Mixed-effects Models (GAMMs) (Wood, 2017) for tongue tip movements and tongue body movements separately. In each of the two models, the dependent variable was the vertical position of the tongue tip/body (i.e., `TonguePosition`).

Our predictors of interest were time (i.e., `Time`), word frequency (i.e., `Freq`), and a factor variable of inflectional status with two levels 'non-inflected' vs. 'inflected' (i.e., `InflStatus`). Word frequency values were obtained for the target words from the SdeWac corpus (Faaß & Eckart, 2013) and log-transformed prior to the analysis. Log-transformed word frequency was distributed approximately in the same range for non-inflected and inflected words (Figure 2). Data points are sparse below log frequency values below 7.

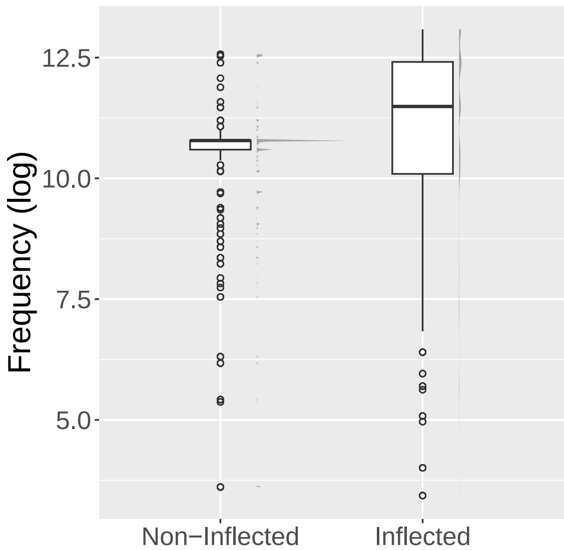


Figure 2: Distributions of log-transformed word frequency for non-inflected and inflected words.

Time was normalized between 0 and 1, corresponding to the onset and the offset of the target vowel [a(ː)]. To compensate for the normalization, the target vowel's duration (i.e., `VowelDuration`) was included as a covariate.

The duration of the target vowel was significantly associated with inflectional condition ($t(368.33) = -4.50$, $p < 0.001$) (Figure 3). The vowel [a(ː)] was significantly longer in duration for inflected words, compared to non-inflected words. The longer duration in inflected words is consistent with previous studies that found similar acoustic lengthening effects in the pre-morphological-boundary condition (Hay, 2007; Li et al., 2020; Plag & Ben Hedia, 2018; Seyfarth et al., 2017; Smith et al., 2012; Song et al., 2013; Strycharczuk & Scobbie, 2016; Sugahara & Turk, 2009).

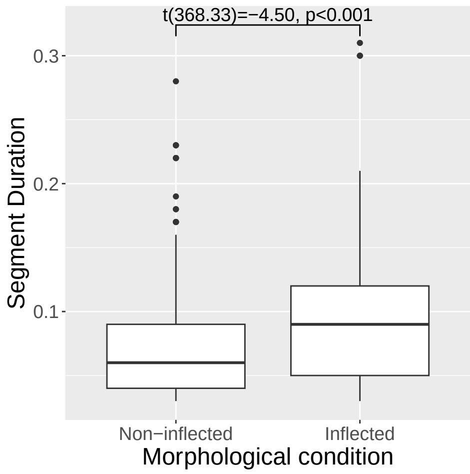


Figure 3: Duration of the target vowel [a(ː)] for inflected and non-inflected words.

In addition, the target vowel's duration was significantly shorter for higher (log) frequency ($r(525) = -0.097$, $p \approx 0.026$), as illustrated in the left panel in Figure 4. The reduction effect of frequency on duration for the present dataset is also in line with previous studies reporting a negative correlation between frequency and segment duration (Aylett & Turk, 2004; Bell et al., 2009; Bell et al., 2002; Gahl, 2008).

Interestingly, separating the inflected and non-inflected words in the present data, high frequency words turned out to be significantly associated with shorter duration ($r(329) = -0.208$, $p < 0.001$) for non-inflected words, but not for inflected words ($r(194) = -0.040$, $p \approx 0.575$) (see the right panel of Figure 4). A linear model regressing segment duration on frequency, inflectional status, and their interaction supports the presence of the interaction ($t(523) = 2.724, p \approx 0.007$). This result suggests that frequency effects play out in different ways for morphologically simple and morphologically complex words.

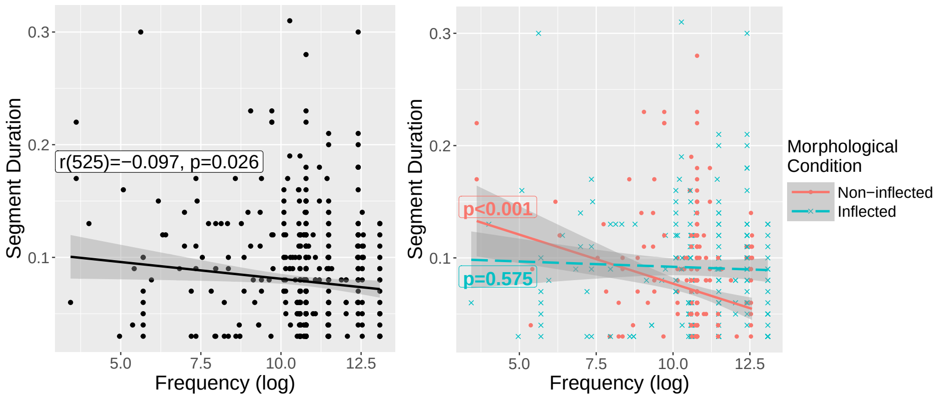


Figure 4: Correlation of frequency with the target vowel's duration, aggregating (left plot) and separating (right plot) the inflectional condition.

With respect to random effects, speaker and word are two common choices in regression modeling. Although there were differences in the number of tokens uttered by the speakers (see Figure 1), including speaker as a random-effect was relatively unproblematic (i.e., `Speaker`). However, as many of the word types were represented by just a single speaker (57%, see Figure 5), inclusion of word as random-effect was not advisable, as it would lead to an over-specified model (see, e.g., Baayen & Linke, 2020).

As the segments preceding and following the target vowel influence the vowel's articulation (Öhman, 1966), we included random effect factors for these two sets of segments (i.e., `PrevSeg` and `NextSeg`). The distributions of the segments before and after the target vowel are illustrated in Figure 6.

Given these predictors, we fitted generalized additive mixed-effects models to the dataset for vertical positions of the tongue tip and the tongue body, using the function `bam` of the package `mgcv` (Wood, 2017) in `R` (R Core Team, 2022). As we are interested in the difference surface of time by frequency for simple and complex words, we adopted a tensor product smooth with 0/1 coding for inflectional status, as follows:

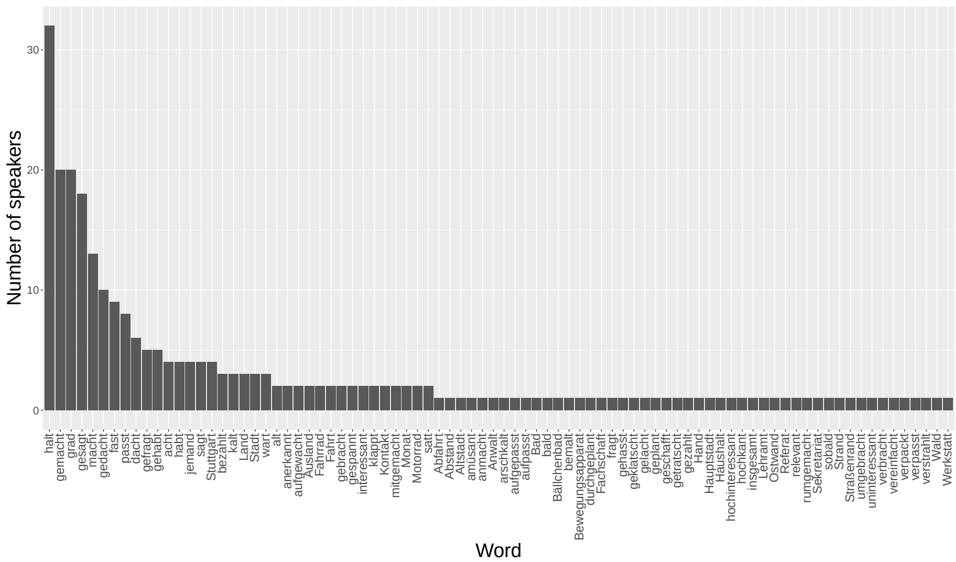


Figure 5: Distribution of the word types across the speakers in the present dataset.

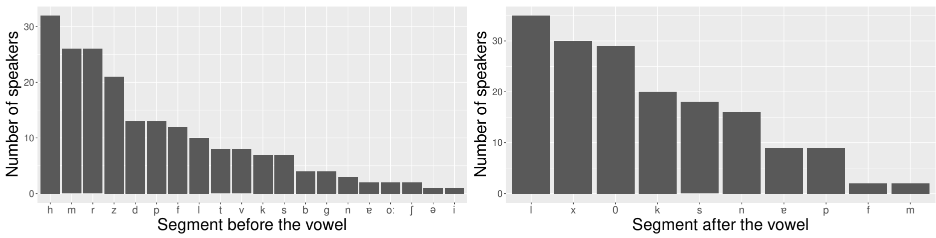


Figure 6: Distributions of the segments before and after the vowel of interest.

```
TonguePosition ~ s(Time, Speaker, bs='fs', k=3, m=1) +
                 s(PrevSeg, bs='re', k=3) +
                 s(NextSeg, bs='re', k=3) +
                 s(VowelDuration, k=3) +
                 ti(VowelDuration, Time, k=c(3,3)) +
                 te(Freq, Time, k=c(3,3)) +
                 te(Freq, Time, by=InflStatus, k=c(3,3)) +
                 InflStatus
```

2.2 Results

2.2.1 Tongue tip

The fitted GAMM for tongue tip positions revealed that articulations of the stem vowel [a(ː)] were significantly lower in general for inflected words, compared to non-inflected words, as indicated by the main effect listed in the second row of the upper part of Table 1 ($\beta = -4.976, p < 0.001$). In addition, the interaction term of inflectional status with `Time` and `Freq` showed that, for our data, the two regression surfaces for non-inflected and inflected words were significantly different, as shown in the last row of the lower part of Table 1.

Table 1: Summary of the model for the tongue tip.

A. Parametric terms	Estimate	Std.Error	*t*-value	*p*-value
Intercept	5.589	2.206	2.533	0.011
Inflected	-4.976	0.467	-10.667	<0.001
B. Smooth terms	**edf**	**Ref.df**	***F***	***p*-value**
s(Time, Speaker)	97.070	104.000	668.328	<0.001
s(PrevSeg)	18.172	19.000	252.770	<0.001
s(NextSeg)	8.265	9.000	1118.025	<0.001
s(VowelDuration)	1.006	1.011	31.340	<0.001
ti(Time, VowelDuration)	1.988	2.002	66.574	<0.001
te(Freq, Time)	7.652	7.929	38.709	<0.001
te(Freq, Time):Inflected	7.510	7.891	15.828	<0.001

This interaction is visualized in Figure 7. The x-axis represents normalized time, and the y-axis log-transformed frequency. The leftmost and middle panels pertain to non-inflected and inflected words respectively. The rightmost panel displays the difference surface for inflected words, namely differences between the surfaces of non-inflected and inflected words. Warmer colors represent higher tongue positions. Since the target vowel is [a(ː)], higher tongue positions indicate articulatory reduction.

For non-inflected words in the leftmost panel, the tongue tip raises as time proceeds, and reaches its highest elevation at the offset of the vowel. This tongue tip raising toward the offset of the stem vowel [a(ː)] is most likely due to anticipatory coarticulation between the stem vowel [a(ː)] and the following word-final [t]. This coarticulatory raising is present irrespective of frequency. The amount of raising, however, depends on frequency: changes in tongue positions over time are greater for lower frequency words than for higher frequency words. Conversely, higher-frequency words are realized with higher tongue positions. This trend is most clearly visible early in the vowel and less so near the end of the vowel. In other words, higher-frequency words have higher and flatter trajectories of the tongue tip throughout the stem vowel [a(ː)].

The difference surface is presented in the rightmost panel of Figure 7. Addition of this difference surface to the surface of the non-inflected words (i.e., the leftmost panel) results in the predicted surface for the inflected words (i.e., the middle panel). The middle panel shows that the reduction effect of frequency is retained to some extent also for inflected words. However, tongue tip trajectories for inflected words are overall lower and have a greater lowering of the tongue at the center of the vowel. In addition, the coarticulatory raising of the tongue tip towards the offset of the vowel, which was observed for non-inflected words, is also attenuated substantially for inflected words.

2.2.2 Tongue body

The tongue body also showed a significant main effect of the inflectional condition ($\beta = -1.559, p < 0.001$), as shown in the upper part of Table 2. Inflectional status also interacted with frequency and time significantly, albeit to a lesser degree compared to the tongue tip model. As can be seen in the left panel of Figure 8, non-inflected words are articulated with higher tongue body positions as frequency increases. These higher tongue positions for higher frequency words are

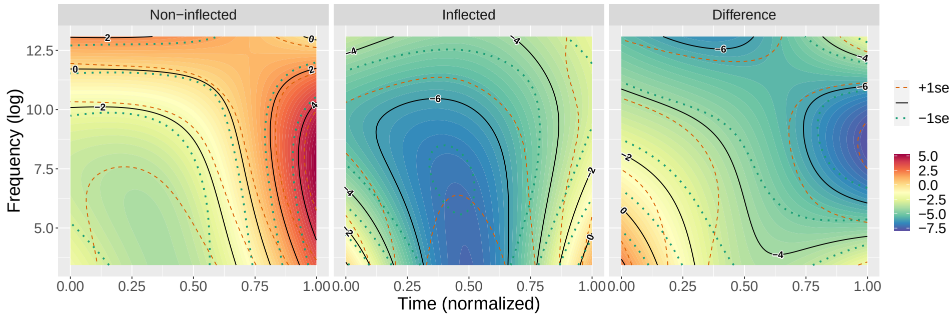


Figure 7: Fitted vertical tongue tip positions as a function of time and frequency for non-inflected words (left), inflected words (middle), and the difference surface (right).

canceled out by the difference surface (the rightmost panel of Figure 8), and as a consequence the tongue trajectories of the tongue body for inflected words (i.e., the middle panel) are minimal, staying relatively low positions.

Table 2: Summary of the model for the tongue body.

A. Parametric terms	Estimate	Std.Error	*t*-value	*p*-value
Intercept	9.387	1.445	6.494	<0.001
Inflected	-1.559	0.410	-3.800	<0.001
B. Smooth terms	**edf**	**Ref.df**	***F***	***p*-value**
s(Time, Speaker)	40.144	104.000	372.145	<0.001
s(PrevSeg)	17.168	19.000	29.081	<0.001
s(NextSeg)	6.798	9.000	111.830	<0.001
s(VowelDuration)	1.003	1.006	0.863	0.355
ti(Time, VowelDuration)	2.905	3.491	5.197	0.001
te(Freq, Time)	4.794	4.987	17.719	<0.001
te(Freq, Time):Inflected	3.006	3.012	3.289	0.020

2.3 Interim summary

For both of the two tongue sensors, the GAMMs revealed higher positions for higher frequency words. In addition, the reduction (tongue-raising) effect of frequency was attenuated for inflected words as compared to non-inflected words.

Overall lower tongue positions for inflected words (the main effects indepen-

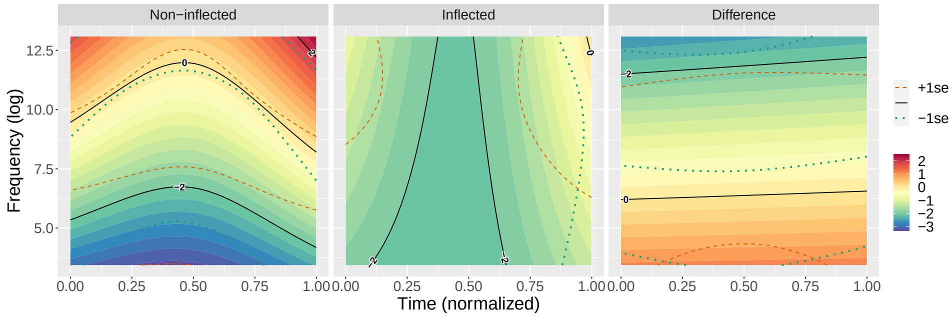


Figure 8: Fitted tongue body height as a function of time and frequency, for non-inflected words (left), inflected words (middle), and the difference surface (right).

dent from frequency and time) are consistent with the paradigm uniformity hypothesis (Seyfarth et al., 2017), according to which phonetic realizations in the pre-morphological-boundary condition should be enhanced. However, this hypothesis does not explain the interaction of the effects of frequency and inflectional status observed for both of the tongue sensors in the current study.

Increases in tongue height hand in hand with increases in frequency reflect articulatory reduction for the stem vowel [a(ː)]. This effect of frequency dovetails well with the smooth signal redundancy hypothesis (Aylett & Turk, 2004), and is consistent with a number of studies that report reduced phonetic realizations (e.g., Gahl, 2008). However, the smooth signal redundancy hypothesis does not predict attenuation of the reduction effect for inflected words. In the current dataset, we observed a much weaker reduction effect of frequency for inflected words. The attenuated reduction effect of frequency may be due to the opposing pressure of enhancing phonetic realizations for clearer articulations. Such opposing enhancement pressure is at least partially in line with the paradigmatic enhancement hypothesis (Kuperman et al., 2007) and the kinematic improvement hypothesis (Tomaschek, Tucker, et al., 2018). However, the absence of such enhancement pressure for non-inflected words remains unaccounted for.

None of these hypotheses fully explain the articulation patterns observed in the present study for inflected and non-inflected words sufficiently. Therefore, in the next section, we investigate whether the observed patterns of articulation can be explained more precisely in terms of words' inflectional semantics.

3 Morphological boundaries or semantics

Several studies framed within the theory of the discriminative lexicon model (Baayen et al., 2019; Chuang et al., 2020; Gahl & Baayen, 2024; Stein & Plag, 2021) have reported phonetic enhancement for word-forms as well as for phonetic segments (Tomaschek et al., 2019; Tomaschek & Ramscar, 2022) that are better-supported by their corresponding semantics. For semantically transparent inflected words, strong links between their forms and meanings are expected, and it is conceivable that these strong links underlie the enhanced articulations reported above.

To test this hypothesis, we will first define a measure of semantic support and show that the semantic measure is correlated with inflectional status. Subsequently, the measure will be used as a predictor for tongue trajectories in GAM regression models.

3.1 Semantic measures derived from the DLM

The discriminative lexicon model (DLM: Baayen et al., 2018; Baayen et al., 2019) is a computational model of lexical processing that works with numerical representations of word-forms and word-semantics. In this study, we represent word-forms with zero/one binary vectors that encode which triphones are present in a word. These vectors are brought together as row vectors of a word-by-triphone matrix (henceforth $\boldsymbol{C}$). Each word vector (row) in $\boldsymbol{C}$ contains 1 where the triphone in question is contained in the word and 0 otherwise.

Word-meanings are represented by word embeddings. We adopted a pre-trained word2vec model (Müller, 2015) which represented word-meanings with 300 dimensional vectors. These vectors are combined as row vectors of a word-by-semantics matrix (henceforth $\boldsymbol{S}$).

We set up the $\boldsymbol{C}$ (64068, 14404) and $\boldsymbol{S}$ (64068, 300) matrices for all those words in the CELEX database (Baayen et al., 1995), whose frequency was greater than 0 and for which pre-trained embeddings were available. The DLM posits simple linear mappings between form and meaning matrices. Given $\boldsymbol{C}$ and $\boldsymbol{S}$, a weight matrix $\boldsymbol{F}$, used for modeling comprehension, can be estimated by solving $\boldsymbol{CF} = \boldsymbol{S}$. The obtained $\boldsymbol{F}$ can then be used to estimate a predict semantic matrix $\hat{\boldsymbol{S}}$ by post-multiplication of $\boldsymbol{C}$ by $\boldsymbol{F}$ (i.e., $\boldsymbol{CF} = \hat{\boldsymbol{S}}$). Rows of $\hat{\boldsymbol{S}}$ represent predicted word meanings. Conceptually, these are the meanings understood by the system given the corresponding word-forms. Similarly, a weight matrix $\boldsymbol{G}$ can be estimated for modelling a part of the speech production process by solving $\boldsymbol{SG} = \boldsymbol{C}$. The estimated $\boldsymbol{G}$ maps $\boldsymbol{S}$ onto $\hat{\boldsymbol{C}}$ (i.e., $\boldsymbol{SG} = \hat{\boldsymbol{C}}$). Rows of $\hat{\boldsymbol{C}}$ are predicted semantic support for word-forms. This method of estimating $\boldsymbol{F}$ and $\boldsymbol{G}$ is called "endstate of learning". For other learning methods implemented for the DLM, see Heitmeier et al. (2022).

Using the endstate-of-learning method in the framework of the discriminative lexicon model, Gahl and Baayen (2024) found that the sum of semantic support from the word's meaning to the triphones constituting the word was predictive for word-duration of English homophones (Gahl & Baayen, 2024). Greater semantic support was associated with longer duration (Gahl & Baayen, 2024). For a word i, the semantic support for triphone j is:

$$\text{SemSup}_{i,j} = \hat{\boldsymbol{C}}_{i,j} \tag{1}$$

Let $\mathscr{C}_i$ a set of triphones constituting a word i. The sum of semantic support for all the component triphones of a word i, which we call `SemSupWord`, is:

$$\text{SemSupWord}_i = \sum_{k \in \mathscr{C}_i} \hat{\boldsymbol{C}}_{i,k} \tag{2}$$

In addition to `SemSupWord`, we considered the triphone centered around the vowel (henceforth the vowel triphone) and the triphone centered around the exponent (henceforth the suffix triphone). Let v and s denote the indices of the vowel and suffix triphones. The semantic support from a word i to its vowel triphone (`SemSupVowel`) and the suffix triphone (`SemSupSuffix`) are defined as

$$\text{SemSupVowel}_i = \hat{\boldsymbol{C}}_{i,v} \tag{3}$$

and

$$\text{SemSupSuffix}_i = \hat{\boldsymbol{C}}_{i,s}. \tag{4}$$

Along with these measures of semantic support, prediction accuracy of the trained LDL model (i.e., `PredAcc`) was also considered. Prediction accuracy was quantified as the correlation of the predicted and observed (gold-standard) row vectors of $\hat{\boldsymbol{C}}$ and $\boldsymbol{C}$. Denoting the i-th row vector of $\hat{\boldsymbol{C}}$ (and $\boldsymbol{C}$) as $\hat{\boldsymbol{C}}_{i,*}$ ($\boldsymbol{C}_{i,*}$), we have:

$$\text{PredAcc}_i = \text{cor}(\hat{\boldsymbol{C}}_{i,*}, \boldsymbol{C}_{i,*}) \tag{5}$$

`PredAcc` was expected to be correlated with the semantic support measures to some extent, especially `SemSupWord`, because well-predicted word-form-vectors should have higher values (only) for their correct component triphones.

In addition, we also defined another measure that focused on uncertainty among predicted form vectors. Uncertainty among predicted forms (i.e., `UncertProd`) is the product of the correlation of the predicted and observed form vectors and the correlation's rank:

$$\text{UncertProd}_i = \sum_k \left(\text{cor}(\hat{\boldsymbol{C}}_{i,*}, \boldsymbol{C}_{k,*}) \times \text{rank}(\text{cor}(\hat{\boldsymbol{C}}_{i,*}, \boldsymbol{C}_{k,*}))\right). \tag{6}$$

The counterpart of this measure for comprehension side is

$$\text{UncertComp}_i = \sum_k \left(\text{cor}(\hat{\boldsymbol{S}}_{i,*}, \boldsymbol{S}_{k,*}) \times \text{rank}(\text{cor}(\hat{\boldsymbol{S}}_{i,*}, \boldsymbol{S}_{k,*}))\right). \tag{7}$$

These uncertainty measures are illustrated in Figure 9. The left panel presents an example of high uncertainty. The shaded part under the curve represents the uncertainty measure. The predicted word with the highest correlation is found at the right hand side of the plot, with the biggest rank, but many other words are also supported by high correlations. Therefore, even if the most strongly supported word is the correct target word, there are also many other words that are "competitive". In contrast, the right panel shows a case of low uncertainty in prediction. Only one of the possible words is strongly supported with a very high correlation coefficient, and the other words are not well supported. In this example, the word with the greatest rank is supported not only well supported, but also there is little uncertainty about what word is the best candidate.

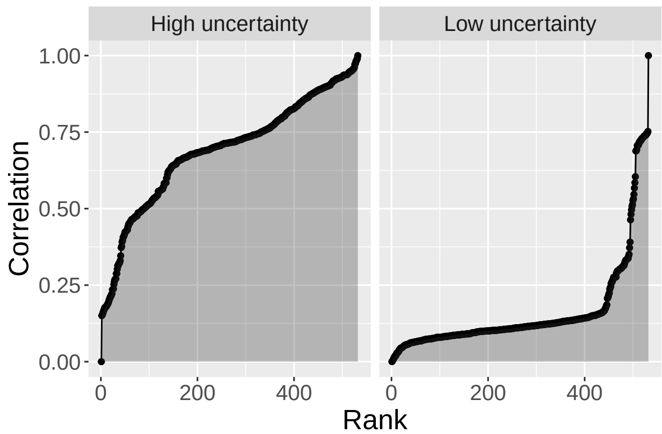


Figure 9: Illustration of high and low uncertainty cases.

In addition, the counterpart of semantic support for the comprehension side of the mappings was also considered. This measure, which we call functional load (i.e., `FuncLoad`), quantifies how much triphones help to identify the target word in the comprehension mapping. The functional load of a triphone is defined as the correlation of that triphone's row vector in $\boldsymbol{F}$ and the semantic vector of its carrier word in $\hat{\boldsymbol{S}}$. The `FuncLoad` of the j-th triphone to the i-th word is given by

$$\text{FuncLoad}_{j,i} = \text{cor}(\boldsymbol{F}_{j,*}, \hat{\boldsymbol{S}}_{i,*}). \tag{8}$$

As for `SemSup`, `FuncLoad` can also be defined for the vowel triphone and the suffix triphone:

$$\text{FuncLoadVowel}_{i} = \text{cor}(\boldsymbol{F}_{v,*}, \hat{\boldsymbol{S}}_{i,*}), \tag{9}$$

$$\text{FuncLoadSuffix}_{i} = \text{cor}(\boldsymbol{F}_{s,*}, \hat{\boldsymbol{S}}_{i,*}). \tag{10}$$

The last measure we considered is the length of a semantic vector (`SemLen`). `SemLen` is simply the L1norm of a semantic vector:

$$\text{SemLen}_{i} = \sum_{j} |S_{ij}|. \tag{11}$$

3.2 Correlation between inflectional status and semantic support

How are these semantic measures related to inflectional status? In this section, we first address this question using variable importance measures based on the Random Forest analysis (Breiman, 2001). Subsequently, we look in more detail into how the most important semantic measures pattern with respect to inflectional status.

To this end, all the words from the CELEX database (Baayen et al., 1995) with the stem vowel [a(ː)] and the word-final segment [t], whose frequency was more than 0, were selected. At most one intervening segment between [a(ː)] and [t] was allowed. The resulting dataset comprised 1392 words. Inflectional status was assigned with help of the inflectional information recorded in CELEX. For example, in CELEX, *macht* [maxt] 'makes' is coded as "3SIE,2PIE,rP". The code stands for "third-person singular indicative present (3SIE)", "second-person plural indicative present (2SPIE)", and "imperative plural (rP)". Appendix A provides a complete list of feature bundles and their classification as either "inflected" or "non-inflected".

3.2.1 Variable importance

Inflectional status was entered as the dependent variable in a Random Forest analysis. The number of predictors being considered for a given subsample (i.e., for each split of the tree) was set to three (of the nine semantic measures introduced above), based on a grid-search using the function `train` of the `caret` package (Kuhn, 2021) in `R` (R Core Team, 2022).

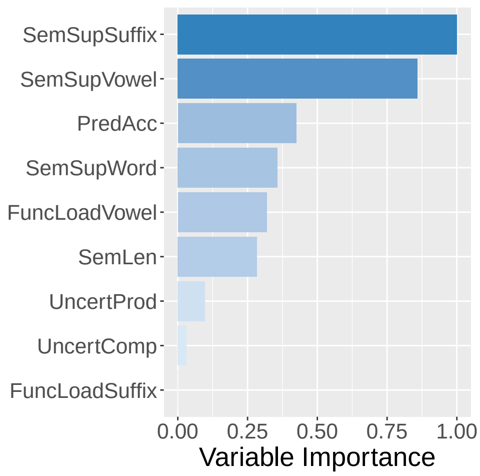


Figure 10: Variable importance of the semantic measures.

The variable importances of the semantic measures are presented in Figure

10. SemSupSuffix is the best supported predictor for inflectional status, followed by SemSupVowel. SemSupWord was not as predictive as SemSupSuffix and SemSupVowel. The good performance of SemSupSuffix fits well with the fact that the exponent *-t* has well-defined inflectional meanings, and thus differs from non-inflectional word-final [t].

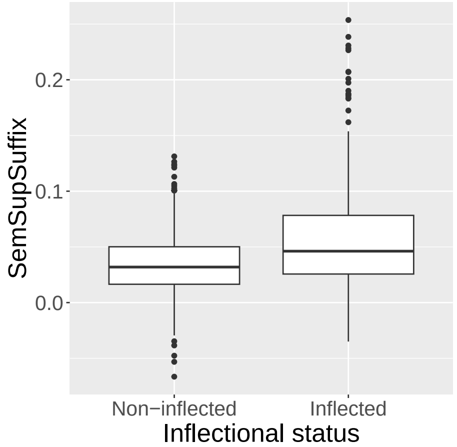
In what follows, we focus on three of the best-supported measures, which are namely SemSupSuffix, SemSupVowel, and SemSupWord, and look into how they are correlated with inflectional status. Since PredAcc is highly correlated with SemSupWord ($r = 0.857$), this predictor was not considered further.

3.2.2 Predicting inflectional status with semantic measures

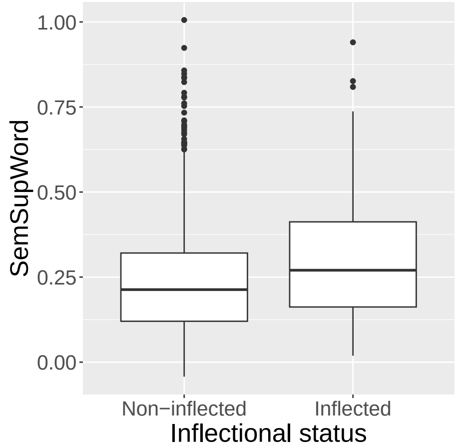
Inflected words had significantly higher values of semantic support for the suffix and the entire word (U=152201, N1=922, N2=470, $p < 0.001$ for SemSupSuffix; U=172134, N1=922, N2=470, $p < 0.001$ for SemSupWord), as illustrated in Figures 11a and 11b respectively. By contrast, inflected words were associated with significantly lower SemSupVowel (U=266563, N1=922, N2=470, $p < 0.001$) as can be seen in Figure 11c.

Subsequently, we fitted logistic regression models, in which the dependent variable was inflectional status. The goal of the logistic models was to predict the probability of a word being inflected. The predictors were SemSupSuffix, SemSupVowel, and SemSupWord. Due to moderate correlations among the three semantic support measures, three logistic regression models were fitted for each of the three semantic measures. Each of the three models showed that the semantic support measures were always highly significant ($p < 0.001$).

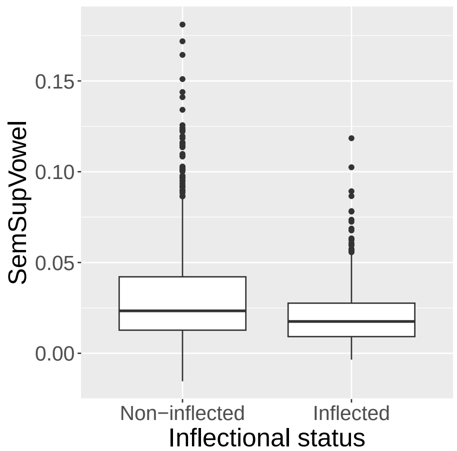
As illustrated in Figure 11d, the effects of the three semantic support measures were qualitatively different. SemSupSuffix was associated the most strongly with the probability of inflectedness. The greater SemSupSuffix becomes, the more



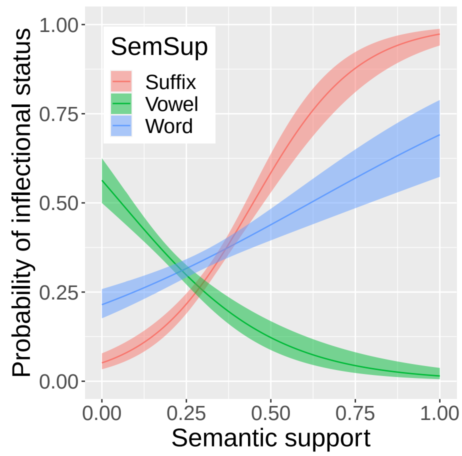
(a) Distributions of `SemSupSuffix` for inflected and non-inflected words.



(b) Distributions of `SemSupWord` for inflected and non-inflected words.



(c) Distributions of `SemSupVowel` for inflected and non-inflected words.



(d) Probability of inflectional status predicted by the three semantic support measures.

Figure 11: Comparison of `SemSupSuffix`, `SemSupVowel`, and `SemSupWord`.

likely the word in question is to be inflected. A similar effect was observed also for `SemSupWord`, albeit to a lesser degree. In contrast, higher `SemSupVowel` was correlated with lower probability of inflectedness.

In line with the results of the variable importances obtained with a Random Forest analysis above, the present analyses confirmed that `SemSupSuffix` was the most effective predictor for inflectional status. Accordingly, in the next section, we focus on `SemSupSuffix` to clarify whether `SemSupSuffix` is also predictive for tongue tip trajectories. Considering that `SemSupSuffix` was greater for inflected words and that inflected words showed articulatory enhancement (Section 2), greater `SemSupSuffix` is expected to be associated with articulatory enhancement. This hypothesis will be tested in the next section. In addition, performance of `SemSupSuffix` will be compared with that of the binary predictor of inflectional status.

3.3 Predicting tongue trajectories from semantics

We used the same dataset as in Section 2 to compare performance of semantic support for suffix (i.e., `SemSupSuffix`) with that of inflectional status as a binary predictor. Some words were not available in CELEX or the pre-trained word2vec model. As a consequence, 5.33% of the data points were lost.

For the remaining data, a GAMM was fitted with the same model structure as in Section 2 except for the predictor of inflectional status. Inflectional status was represented by a binary factor in Section 2 in interaction with normalized time and log-transformed frequency. In the following analyses, the binary factor was replaced with `SemSupSuffix`. We fitted the following model to the data, again including the three-way interaction:

```
TonguePosition ~ s(Time, Speaker, bs='fs', k=3, m=1) +
                 s(PrevSeg, bs='re', k=3) +
                 s(NextSeg, bs='re', k=3) +
                 s(VowelDuration, k=3) +
                 ti(VowelDuration, Time, k=c(3,3)) +
                 te(Time, SemSupSuffix, Freq, k=c(3, 3, 3))
```

The model with `SemSupSuffix` required one less edf, and nevertheless improved the model fit significantly by 142.87 AIC units (by 62.64 ML scores), compared to the model with a binary factor of inflectional status[1]. All the terms in the model were well-supported (except for the intercept; see Table 3). Figure 12 illustrates the interaction of `SemSupSuffix` by frequency at the center of the vowel. In this figure, the x-axis represents frequency, and the y-axis `SemSupSuffix`. Warmer colors represent higher tongue tip positions. Since the target vowel is [a(ː)], higher tongue positions correspond to articulatory reduction.

Table 3: Summary of the model with `SemSupSuffix`.

A. Parametric terms	Estimate	Std.Error	*t*-value	*p*-value
Intercept	2.813	2.564	1.097	0.273
B. Smooth terms	edf	Ref.df	*F*	*p*-value
s(Time, Speaker)	97.853	104.000	602.212	<0.001
s(PrevSeg)	16.671	17.000	544.862	<0.001
s(NextSeg)	7.870	8.000	2177.474	<0.001
s(VowelDuration)	1.017	1.033	18.459	<0.001
ti(Time, VowelDuration)	3.631	3.909	29.393	<0.001
te(Time, SemSupSuffix, Freq)	20.312	22.235	30.917	<0.001

Figure 12 shows that higher `SemSupSuffix` goes hand in hand with lower tongue trajectories for higher frequency, indicating that the enhancement effect of `SemSupSuffix` is limited to higher frequency words. From the perspective of frequency effects, higher frequency is associated with higher tongue positions for low `SemSupSuffix` values, indicating articulatory reduction effects of frequency. In contrast, when `SemSupSuffix` is high, an increase in frequency is tied with lowering of the tongue tip, indicating articulatory enhancement. Since greater `SemSupSuffix` is correlated with inflectedness (Section 3.2.2), the current result is in line with the strong and attenuated reduction effects of frequency for non-inflected and inflected words respectively, reported in Section 2.

[1]No model comparison test was necessary because the model with `SemSupSuffix` was simpler and better than the model with a binary factor variable of inflectional status.

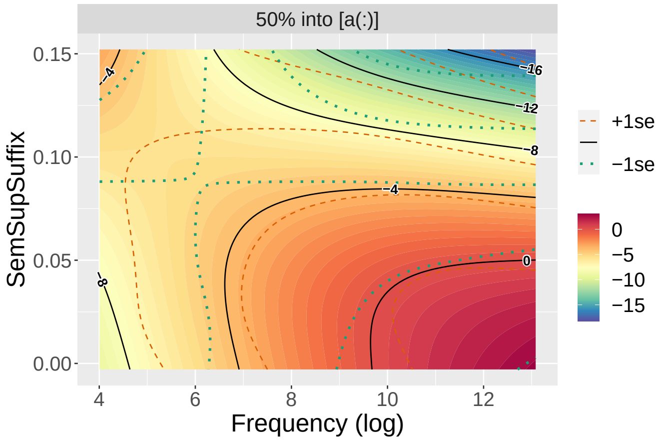


Figure 12: Tongue tip height as a function of frequency and `SemSupSuffix` at the middle of the vowel. Warmer colors represent high and colder colors represent low positions.

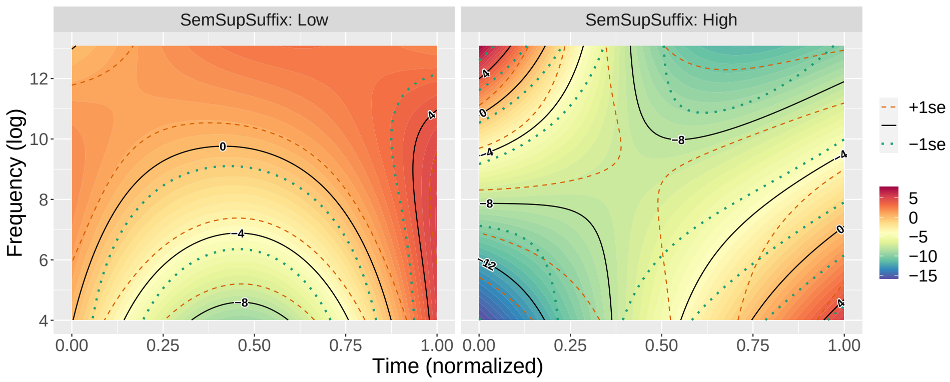


Figure 13: Tongue tip height as a function of time and frequency. `SemSupSuffix` is discretized to low and high values, corresponding to 0.01 and 0.99 quantiles. Warmer colors represent high and colder colors represents low positions.

Figure 12 does not show in details how tongue trajectories over time are modulated by frequency and `SemSupSuffix`, because thus far time was fixed at the middle of the vowel. Figure 13 zooms in on time, illustrating qualitative differences in tongue trajectories as a function of time, frequency, and `SemSupSuffix`. For illustration, `SemSupSuffix` is discretized into high and low values, corresponding to 1% and 99% quantiles.

When `SemSupSuffix` is low (in the left panel), lower frequency is associated with lower tongue trajectories, and higher frequency is associated with higher tongue trajectories. On the other hand, when `SemSupSuffix` is high (in the right panel), high frequency words are articulated with tongue trajectories with a greater lowering from the middle to the offset of the vowel.

3.4 Interim summary

In this section, we showed that semantic support from word-meanings to word-final triphones (i.e. `SemSupSuffix`) outperformed other semantic measures in variable importance estimated by a Random Forest (Breiman, 2001). In line with this analysis with variable importance, `SemSupSuffix` also predicted inflectional status the most effectively. Higher `SemSupSuffix` was associated with higher probability of words being inflected, and it was also interacted with frequency. Because of the interaction with `SemSupSuffix`, higher frequency was associated with higher tongue trajectories, indicating reduced articulations, for low `SemSupSuffix`, while higher frequency was correlated with lower tongue positions, indicating enhanced articulations, for high `SemSupSuffix`. These observed patterns are in line with the patterns observed in Section 2. In Section 2, higher frequency was associated with strong articulatory reduction for non-inflected words, while the reduction effect was attenuated for inflected words.

Thus far in the current study, we have mainly focused on the semantic support for the word-final triphone, which is centered around the exponent [t]. We also considered a model in which the semantic support for the triphone straddling the vowel (i.e., `SemSupVowel`) was considered instead. This model showed that a greater semantic support for the vowel leads to a lower position of the tongue tip. At the same time, higher word frequency predicted higher tongue positions, irrespective of the amount of semantic support for the vowel (see Appendix C for detail). As only 14 out of the 70 word types in the current dataset had a stem that ended in a vowel, a vast majority of the target words had a vowel triphone that did not include the inflectional exponent [t]. From these observations, we conclude that on the one hand, greater semantic support for the vowel gives rise to enhanced articulation of the stem vowel, but that the effect of frequency works against this, giving rise to higher tongue positions.

4 Discussion

In what follows, we first explore possible explanations for the observed patterns. Subsequently, we propose our interpretation and lay out implications of our findings for existing theories.

Higher frequency has been reported to be correlated with both phonetic reduction (e.g. Aylett & Turk, 2004) as well as enhancement (e.g. Kuperman et al., 2007; Tomaschek, Tucker, et al., 2018; Tomaschek et al., 2021). These seemingly contradictory effects may be due to morphological status of the items being investigated. When the reduction effect is observed, morphologically simple words are always included. On the other hand, the enhancement effect has been observed only for morphologically complex words.

In order to clarify the role of morphological structure, we focused on inflected and non-inflected words in German. The target words shared the same rhyme structure with the stem vowel being [a(ː)] and with the word-final segment being [t]. The word-final [t] was a part of the stem for non-inflected words, while it was an exponent for inflected words. Vertical tongue tip and body positions were fitted with Generalized Additive Mixed-effects Models (GAMMs) (Wood, 2017) as a function of time, frequency, and inflectional status together with random effect factors and control covariates.

The tongue tip and body models both showed significant effects of inflectional status. Inflected words showed lower tongue tip/body positions on average than non-inflected words. Since the vowel [a(:)] was investigated and followed by a morphological boundary in inflected words, these results suggest enhanced articulatory realizations in the pre-morphological-boundary condition.

Pre-morphological boundary enhancement is in harmony with the paradigm uniformity hypothesis (Seyfarth et al., 2017), which predicts that members of the same paradigm become similar in phonetic realizations to each other. However, the quality and degrees of articulatory enhancement were significantly modulated by frequency in the current study. Inflected words retained lower tongue positions, namely more enhanced tongue positions, compared to non-inflected words, as frequency increased. This interaction of inflectional status and frequency was observed for the tongue tip and the tongue body both.

Increased degrees of articulatory enhancement (implying decreased degrees of articulatory reduction) for higher frequency inflected words are consistent with the articulatory improvement hypothesis (Tomaschek, Tucker, et al., 2018). Higher frequency words are articulatorily well-practiced and therefore their articulations are faster and more enhanced. However, under this hypothesis, not only inflected words but also non-inflected words should be enhanced with increasing frequency. This, however, was not the case in the present study.

In the present study, we observed that non-inflected words were realized with greater degrees of articulatory reduction as frequency increases. This reduction effect is in line with the smooth signal redundancy hypothesis (Aylett & Turk, 2004). Higher frequency can go hand in hand with higher redundancy and lower amounts of information (surprisal). According to Aylett and Turk (2004), this motivates articulatory reduction. However, the smooth signal redundancy hypothesis does not take into consideration the morphological status of the word in question. Consequently, the hypothesis predicts the same degree of phonetic reduction also for inflected words, which was not the case in the present study.

Why do inflected words show less degrees of reduction, while non-inflected words show a strong reduction effect, as frequency increases? One systematic difference between inflected and non-inflected words is the presence and absence of inflectional meanings. In German, inflectional meanings are mostly expressed by and tied with inflectional suffixes. Strong form-meaning relations have been found to be a source of phonetic enhancement: Gahl and Baayen (2024) report that semantically better-supported words are realized with longer durations; Tomaschek et al. (2019) and Tomaschek and Ramscar (2022) report that semantically better-supported segments are realized with longer durations and more peripheral vowel formants. This suggests that inflectional meanings may provide good semantic support for their corresponding inflectional suffixes, which in turn may lead to enhanced realizations in the corresponding inflected words.

This hypothesis was addressed, using the discriminative lexicon model (DLM: Baayen et al., 2018; Baayen et al., 2019). Computational modeling revealed that the semantic support for the word-final triphone (`SemSupSuffix`) outperformed other semantic measures such as semantic support for the stem triphone. Greater `SemSupSuffix` was strongly associated with higher probability of inflectedness. Therefore, `SemSupSuffix` can be understood as a continuous counterpart of a categorical factor specifying inflectional status. Replacing the categorical predictor 'inflectional status' by `SemSupSuffix` resulted in a significant improvement in model fit.

`SemSupSuffix` was also shown to be predictive for vertical positions of the tongue tip. For higher-frequency words, a higher `SemSupSuffix` predicted a lower tongue position. From the perspective of frequency effects, `SemSupSuffix` emerged as a modulation of frequency effects. When `SemSupSuffix` was high, high frequency words were articulated with lower tongue positions. When

SemSupSuffix was low, high frequency words were articulated with higher tongue positions. Since high SemSupSuffix was associated with inflected words, the modulation by SemSupSuffix explains why inflected words were less reduced, while non-inflected words showed strong reduction, as frequency increased.

Importantly, this explanation does not require the theoretical concept of a 'morphological boundary'. The present results therefore challenge the classical view of the speech production process such as formalized in the WEAVER++ model (Levelt et al., 1999; Levelt & Wheeldon, 1994; Roelofs, 1997), which operates on morphemes with at least one intermediate symbolic layer between semantics and phonetics. On the other hand, the present results support the hypothesis that better mappings between inflectional meanings and forms (inflectional suffixes) go hand in hand with enhanced realizations (Gahl & Baayen, 2024; Tomaschek et al., 2019; Tomaschek & Ramscar, 2022).

In the DLM model, support from word-meanings for the final [t] is predicted to be much stronger for inflected words, which we have shown to be due to the inflectional semantics that are realized by this exponent. In the current study, the enhancement effect of semantic support was observed for the stem vowel. This strong enhancement of the vowel is likely to be due to coarticulation between the stem vowel and the suffix. This possibility is also supported by greater degrees of modulation of tongue trajectories by semantic support and frequency for tongue tip positions than for tongue body positions (compare Figure 13 in Section 3.3 and Figure 17 in Appendix C). Since the present study investigated [a(ː)] followed by the alveolar exponent [t], it makes sense that the co-articulation with [a(ː)] was more prominent for the tongue tip than the tongue body.

We observed that higher semantic support for the vowel triphone predicted lower positions for the tongue tip. However, the vowel triphone does not include the inflectional exponent. Unlike the final triphone, the vowel triphone is not systematically connected with the inflectional semantics of the [t] exponent. This may explain why, in a model replacing the final triphone with the triphone of the vowel, greater word frequency predicted higher positions of the tongue tip. It is only for the final triphone, and its co-articulatory entanglement with the preceding vowel, that the practice effect of frequency is visible.

Enhancement from semantic support is clearly not the only factor that co-determines articulation. For non-inflected words, greater frequency goes hand in hand with higher tongue positions, which fits well with the argument of Aylett and Turk (2004) that less informative words reduce. For the inflected words in our dataset, we observed attenuated degrees of the reduction effect. This is likely due to the reduction effect being counterbalanced by the articulatory strengthening induced by inflectional semantics.

It is possible to explain the reduction effect of predictability in the framework of the discriminative lexicon model. The present study showed that higher `SemSupSuffix`, namely higher $\hat{\boldsymbol{C}}_{i,s}$, was correlated with phonetic enhancement. On the other hand, greater amount of information is said to also go hand in hand with enhanced realizations. Therefore, the effect of informativity can be integrated as a parameter modifying the strength of a semantic vector. Denoting the amount of information of a word ω at a point in a discourse k by $h_{\omega,k}$, the composite effect of informativity and semantic support can be expressed as $h_{\omega,k}\hat{\boldsymbol{C}}_{i,s}$ (see also Gahl & Baayen, 2024).

In summary, the present study shows how the paradox of two seemingly contradictory frequency effects can be resolved. Frequency effects can show up as different degrees of phonetic reduction, depending on morphological status. For inflected words, what looks like an attenuated reduction effect (and even a clear enhancement effect for some previous studies) is actually a composite of a reduction effect due to lack of informativity (e.g., Aylett & Turk, 2004) and a strengthening effect that is determined by the amount of semantic support that a word's form receives. For inflected words, this amount of support, especially for a word's final triphone, is driven by a word's inflectional semantics. In other words, what would seem to be an effect at the level of word form — a morphological boundary effect — actually is driven by inflectional semantics.

Acknowledgments

The authors would like to thank all the members of the research group FOR2373 Spoken Morphology (DFG grant number 273681275) and especially the project's leading investigator Prof. Ingo Plag. This study was constantly supported and improved by presentations and discussions of this research group and would not have been possible without their supports.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [grant number 273681275].

Supplementary materials

The data and the scripts can be found in https://osf.io/94u6n/

References

Arnold, D., & Tomaschek, F. (2016). The Karl Eberhards Corpus of spontaneously spoken southern German in dialogues — audio and articulatory recordings.

Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum, 9–11.

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349–371. https://doi.org/10.1177/0023830913484891

Aylett, M., & Turk, A. (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. https://doi.org/10.1177/00238309040470010201

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

Baayen, R. H., Chuang, Y.-Y., & Blevins, J. P. (2018). Inflectional morphology with linear mappings. *The Mental Lexicon*, *13*(2), 230–268. https://doi.org/10.1075/ml.18010.baa

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, 1–39. https://doi.org/10.1155/2019/4895891

Baayen, R. H., & Linke, M. (2020). An introduction to the Generalized Additive Model. In M. Paquot & S. T. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 563–591). Springer.

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, *30*(11), 1174–1220. https://doi.org/10.1080/02687038.2016.1147767

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2). *Published by the Linguistic Data Consortium, University of Pennsylvania*.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111. https://doi.org/10.1016/j.jml.2008.06.003

Bell, A., Gregory, M. L., Brenier, J. M., Jurafsky, D., Ikeno, A., & Girand, C. (2002). Which predictability measures affect content word durations? *Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology (PMLA), ISCA Tutorial and Research Workshop (ITRW)*, 1–5.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Chuang, Y.-Y., Lõo, K., Blevins, J. P., & Baayen, R. H. (2020). Estonian case inflection made simple: A case study in Word and Paradigm morphology with linear discriminative learning. In L. Körtvélyessy & P. Štekauer (Eds.), *Complex words: Advances in morphology* (pp. 119–141). Cambridge University Press.

Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement: Contextual and paradigmatic effects on morpheme pronunciation. *Morphology*, *24*(4), 291–323. https://doi.org/10.1007/s11525-014-9243-y

Dinkin, A. J. (2008). The real effect of word frequency on phonetic variation. *Proceedings of the 31st Annual Penn Linguistics Colloquium*, *14*(1), 97–106.

Faaß, G., & Eckart, K. (2013). SdeWaC: A corpus of parsable sentences from the web. In I. Gurevych, C. Biemann, & T. Zesch (Eds.), *Language processing and knowledge in the web* (pp. 61–68). Springer.

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

Gahl, S., & Baayen, R. H. (2024). Time and thyme again: Connecting spoken word duration to models of the mental lexicon. [Manuscript, University of Berkeley and University of Tübingen, accepted for publication in Language].

Hay, J. (2007). The phonetics of 'un'. In J. Munat (Ed.), *Lexical creativity, texts and contexts* (pp. 39–57). John Benjamins.

Heitmeier, M., Chuang, Y., Axen, S., & Baayen, R. H. (2022). Frequency-informed linear discriminative learning. *Proceedings of a workshop in honor of Ingo Plag's 60-th birthday*.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). John Benjamins.

Kuhn, M. (2021). caret: Classification and regression training. https://cran.r-project.org/package=caret

Kuperman, V., Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America*, *121*(4), 2261–2271. https://doi.org/10.1121/1.2537393

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

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239–269.

Li, V. G., Oh, S., Chopra, G., Celli, J., & Shaw, J. A. (2020). Articulatory correlates of morpheme boundaries: Preliminary evidence from intra- and intergestural timing in the articulation of the English past tense. *Proceedings of ISSP 2020 - 12th International Seminar on Speech Production*.

Lin, S. S., Beddor, P. S., & Coetzee, A. W. (2011). Gestural reduction and sound change: An ultrasound study. *ICPhS XVII: Proceedings of the 17th International Congress of Phonetic Sciences*, 1250–1253.

Müller, A. (2015). *Analyse von Wort-Vektoren deutscher Textkorpora* (Bachelor's Thesis). Technische Universität Berlin. https://devmount.github.io/GermanWordEmbeddings

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, *39*, 151–168. https://doi.org/10.1121/1.1909864

Plag, I., & Ben Hedia, S. (2018). The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration. In S. Arndt-Lappe, A. Braun, C. Moulin, & E. Winter-Froemel (Eds.), *Expanding the lexicon: Linguistic innovation, morphological productivity, and ludicity* (pp. 93–116). De Gruyter. https://doi.org/10.1515/9783110501933-095

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

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005b). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, *118*(4), 2561–2569. https://doi.org/10.1121/1.2011150

R Core Team. (2022). R: A language and environment for statistical computing. https://www.r-project.org/

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., & Malouf, R. (2017). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience*, *33*(1), 32–49. https://doi.org/10.1080/23273798.2017.1359634

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

Song, J. Y., Demuth, K., Shattuck-Hufnagel, S., & Ménard, L. (2013). The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound. *Journal of Phonetics*, *41*(3-4), 281–295.

Stein, S. D., & Plag, I. (2021). Morpho-phonetic effects in speech production: Modeling the acoustic duration of English derived words with linear discriminative learning. *Frontiers in Psychology*, *12*(678712). https://doi.org/10.3389/fpsyg.2021.678712

Strycharczuk, P., & Scobbie, J. M. (2016). Gradual or abrupt? The phonetic path to morphologisation. *Journal of Phonetics*, *59*, 76–91. https://doi.org/10.1016/j.wocn.2016.09.003

Sugahara, M., & Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phonology*, *26*(3), 477–524.

Tomaschek, F., Arnold, D., Bröker, F., & Baayen, R. H. (2018). Lexical frequency co-determines the speed-curvature relation in articulation. *Journal of Phonetics*, *68*, 103–116.

Tomaschek, F., Plag, I., Ernestus, M., & Baayen, R. H. (2019). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning [Publisher: Cambridge University Press]. *Journal of Linguistics*, *57*(1), 123–161. https://doi.org/https://doi.org/10.1017/S0022226719000203

Tomaschek, F., & Ramscar, M. (2022). Understanding the Phonetic Characteristics of Speech Under Uncertainty—Implications of the Representation of Linguistic Knowledge in Learning and Processing. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.754395

Tomaschek, F., Tucker, B. V., Fasiolo, M., & Baayen, R. H. (2018). Practice makes perfect: the consequences of lexical proficiency for articulation. *Linguistics Vanguard*, *4*(s2), 1–13.

Tomaschek, F., Tucker, B. V., Ramscar, M., & Baayen, R. H. (2021). Paradigmatic enhancement of stem vowels in regular English inflected verb forms. *Morphology*, *31*(2), 171–199. https://doi.org/10.1007/s11525-021-09374-w

Tomaschek, F., Wieling, M., Arnold, D., & Baayen, H. (2013). Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience. *Proceedings of the 14th Annual Conference of the International Speech Communication Association (INTERSPEECH 2013)*, 1302–1306.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd). CRC Press.

Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology vi* (pp. 75–87). Cambridge University Press. https://doi.org/10.1017/cbo9780511486425.005

Appendices

A Assignment of inflectional status

CELEX tag	Example	Present study
0	*jemand*	non-inflected
13SIA	*stand*	non-inflected
2PIE,rP	*fangt*	inflected
2SIE,3SIE,2PIE	*aufpasst*	inflected
2SIE,3SIE,2PIE,rP	*kratzt*	inflected
2SIE,3SIE,2PIE,rP,pA	*erfasst*	inflected
3SIE,2PIE	*ausmacht*	inflected
3SIE,2PIE,pA	*ausbezahlt*	inflected
3SIE,2PIE,rP	*macht*	inflected
3SIE,2PIE,rP,pA	*bezahlt*	inflected
nP,gP,dP,aP,nS,dS,aS	*Watt*	non-inflected
nS	*Kandidat*	non-inflected
nS,dS,aS	*Land*	non-inflected
nS,dS,aS,nP,gP,dP,aP	*England*	non-inflected
nS,gS,dS,aS	*Hand*	non-inflected
pA	*gemacht*	inflected
pA,3SIE,2PIE,rP	*bestrahlt*	inflected
X	*bald*	non-inflected

B `SemSupSuffix` model for tongue body positions

A GAMM with the same structure as in Section 3.3 (the `SemSupSuffix` model for the tongue tip) was also fitted to vertical tongue body positions. The interaction among time, `SemSupSuffix`, and frequency was supported as shown in the last row of Table 4 below.

Table 4: Summary of the model with `SemSupSuffix` for tongue body positions.

A. Parametric terms	Estimate	Std.Error	*t*-value	*p*-value
Intercept	8.436	1.546	5.455	<0.001
B. Smooth terms	**edf**	**Ref.df**	***F***	***p*-value**
s(Time, Speaker)	57.261	104.000	386.911	<0.001
s(PrevSeg)	15.883	17.000	106.811	<0.001
s(NextSeg)	6.823	8.000	343.344	<0.001
s(VowelDuration)	1.006	1.012	0.181	0.675
ti(Time, VowelDuration)	3.041	3.569	5.663	<0.001
te(Time, SemSupSuffix, Freq)	13.862	15.037	9.612	<0.001

A visualization of the interaction between frequency and `SemSupSuffix` at the center of the vowel (Figure 14) indicates that their effects are minimal in most combinations of values of `SemSupSuffix` and frequency. Patterns of tongue body trajectories are comparable for low and high values of `SemSupSuffix`, while higher frequency is constantly associated with higher tongue body positions.

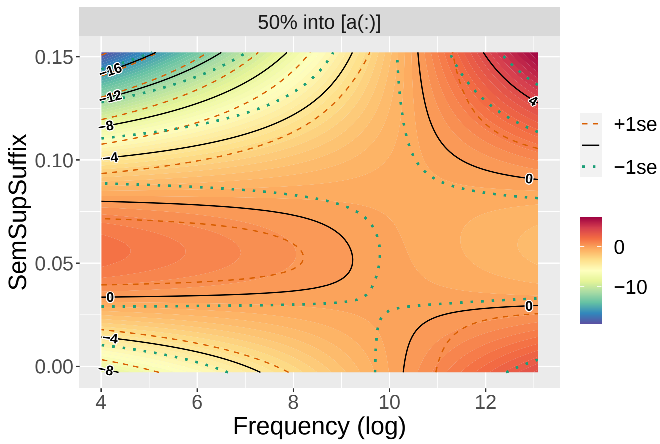


Figure 14: Tongue body height as a function of frequency and `SemSupSuffix` at the middle of the vowel. Warmer colors represent high and colder colors represent low positions.

Tongue body trajectories are also predicted to be very similar for different values of `SemSupSuffix`. Figure 15 displays predicted tongue body trajectories for a large value and a small value of `SemSupSuffix` in the left and right panels respectively. In both of the panels, lower frequency is associated with lower tongue body trajectories, and higher frequency is associated with higher tongue body trajectories.

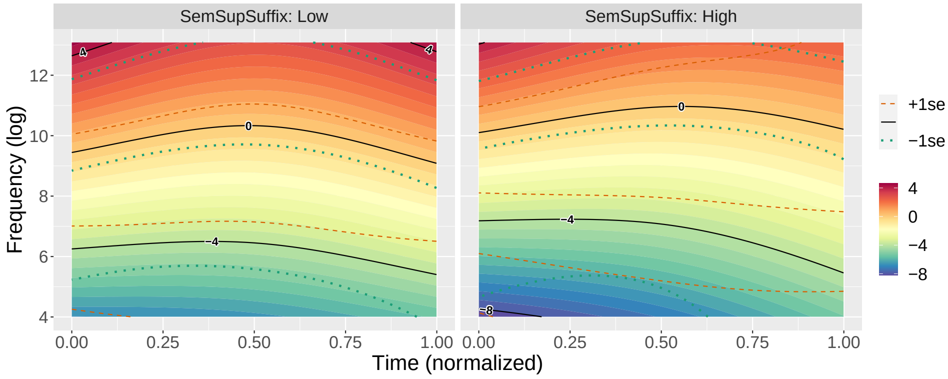


Figure 15: Tongue body height as a function of time and frequency. Semantic support for suffixes (i.e., `SemSupSuffix`) is discretized to low and high values, corresponding to 0.01 and 0.99 quantiles. Warmer colors represent high and colder colors represent low positions.

The current dataset consists of the words with the stem vowel [a(ː)] and the word-final segment [t], which are expected to induce coarticulatory movements mainly for the tongue tip, leaving the tongue body being moved only passively. Therefore, these results suggest that strengthening effects by semantic support mainly influence coarticulatory movements of the tongue.

C `SemSupVowel` models

`SemSupVowel` was distributed in a right-skewed manner. Therefore, the variable was log-transformed in prior to fitting GAMMs. After the log-transformation, `SemSupVowel` was fitted with GAMMs to predict tongue tip and body positions with other control variables and random effects in the same model structure as for `SemSupSuffix` (see Section 3.3 for the model structure), except for replacing

SemSupSuffix for SemSupVowel.

C.1 Tongue tip

The fitted GAMM showed that higher SemSupVowel was constantly associated with lower tongue tip positions, namely clearer articulations (see Table 5 and Figure 16). Figure 17 further illustrates that higher frequency words are articulated with higher and flatter tongue tip trajectories when SemSupVowel is low, while higher frequency words show attenuated degrees of reduction (i.e., tongue-raising effects) when SemSupVowel is high.

Table 5: Summary of the model with log-transformed SemSupVowel for tongue tip positions.

A. Parametric terms	Estimate	Std.Error	*t*-value	*p*-value
Intercept	2.863	2.665	1.074	0.283
B. Smooth terms	**edf**	**Ref.df**	***F***	***p*-value**
s(Time, Speaker)	98.276	104.000	317.487	$<$0.001
s(PrevSeg)	13.718	14.000	507.555	$<$0.001
s(NextSeg)	7.801	8.000	1053.522	$<$0.001
s(VowelDuration)	1.793	1.955	12.648	$<$0.001
ti(Time, VowelDuration)	3.498	3.851	25.312	$<$0.001
te(Time, SemSupVowel, Freq)	24.158	25.262	34.214	$<$0.001

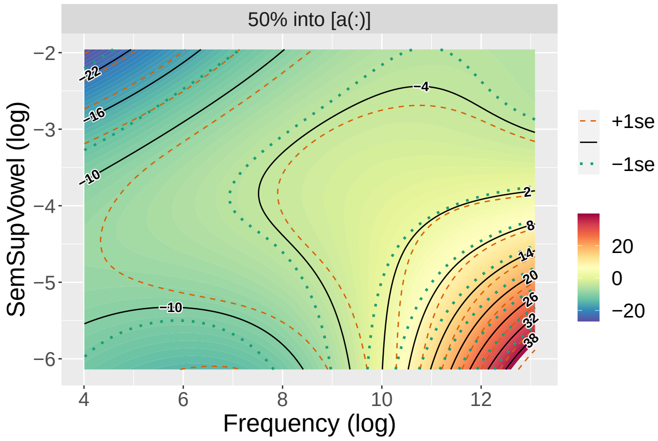


Figure 16: Tongue tip height as a function of frequency and log-transformed SemSupVowel at the middle of the vowel. Warmer colors represent high and colder colors represent low positions.

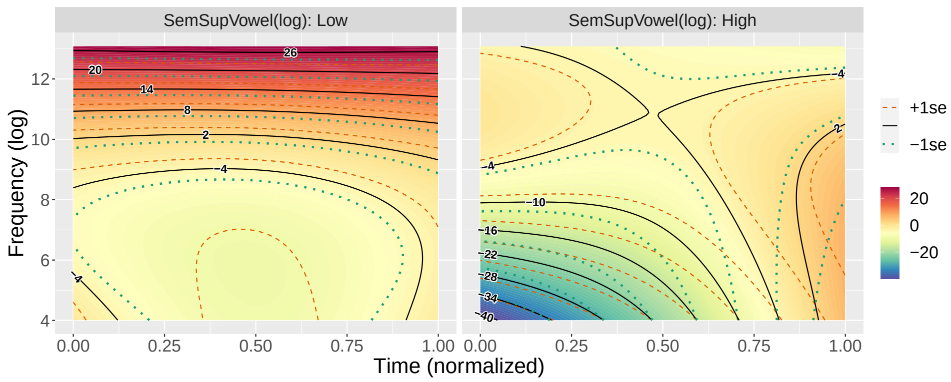


Figure 17: Tongue tip height as a function of time and frequency. SemSupVowel is log-transformed and discretized to low and high values, corresponding to 0.01 and 0.99 quantiles. Warmer colors represent high and colder colors represents low positions.

C.2 Tongue body

The same structure of a GAMM was fitted for tongue body positions (Table 6). The interaction of frequency and SemSupVowel turned out to be a U-shaped effect (Figure 18). This effect is likely due to extreme values predicted for very high and very low SemSupVowel values. For middle values of SemSupVowel, predicted tongue body height is almost always zero, indicating no substantial effect of frequency and `SemSupVowel` is visible in the region. In line with this observation, tongue trajectories are predicted to stay slightly higher than the occlusal plane (i.e. 0) with not much raising or lowering during the vowel, regardless of values of frequency. A possible exception could be tongue body positions at the onset of the vowel for low frequency words with high `SemSupVowel`, where low positions are predicted. However, these predictions are not very reliable due to sparseness of data points below (log) frequency being 7.

Table 6: Summary of the model with log-transformed `SemSupVowel` for tongue body positions.

A. Parametric terms	Estimate	Std.Error	*t*-value	*p*-value
Intercept	8.313	1.419	5.858	<0.001
B. Smooth terms	**edf**	**Ref.df**	***F***	***p*-value**
s(Time, Speaker)	94.481	104.000	173.992	<0.001
s(PrevSeg)	11.967	14.000	44.595	<0.001
s(NextSeg)	6.081	8.000	109.332	<0.001
s(VowelDuration)	1.007	1.014	4.879	0.026
ti(Time, VowelDuration)	1.998	2.017	48.605	<0.001
te(Time, SemSupVowel, Freq)	23.823	25.147	25.460	<0.001

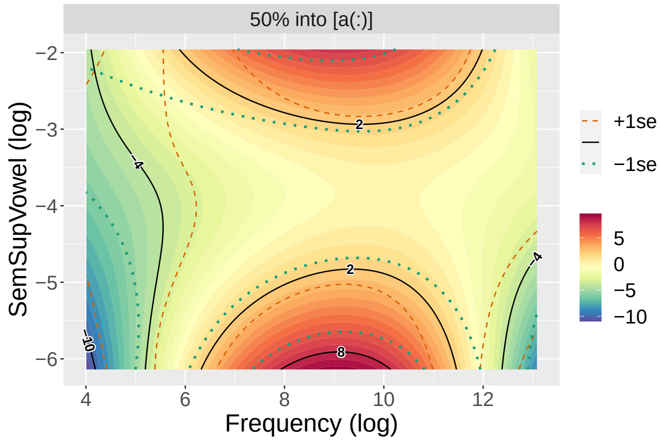


Figure 18: Tongue body height as a function of frequency and log-transformed `SemSupVowel` at the middle of the vowel. Warmer colors represent high and colder colors represent low positions.

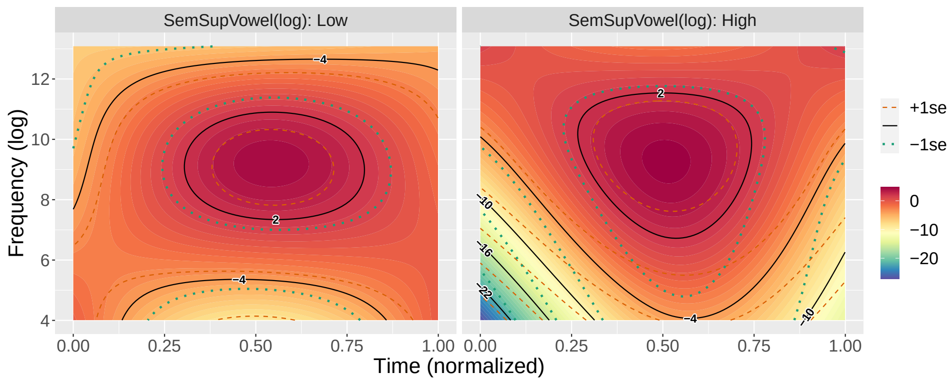


Figure 19: Tongue body height as a function of time and frequency. Semantic support for the stem vowel (i.e., `SemSupVowel`) is log-transformed and discretized to low and high values, which correspond to 0.01 and 0.99 quantiles. Warmer colors represent high and colder colors represents low positions.