Phonetic reduction and paradigm uniformity effects in spontaneous speech

U. Marie Engemann; Heinrich-Heine-Universität Düsseldorf; marie.engemann@hhu.de
Ingo Plag; Heinrich-Heine-Universität Düsseldorf; ingo.plag@hhu.de

Abstract

Recent work on the acoustic properties of complex words has found that morphological information may influence the phonetic properties of words, e.g. acoustic duration. Paradigm uniformity has been proposed as one mechanism that may cause such effects. In a recent experimental study Seyfarth et al. (2017) found that the stems of English inflected words (e.g. *frees*) have a longer duration than the same string of segments in a homophonous mono-morphemic word (e.g. *freeze*), due to the co-activation of the longer articulatory gesture of the bare stem (e.g. *free*). However, not all effects predicted by paradigm uniformity were found in that study, and the role of frequency-related phonetic reduction remained inconclusive. The present paper tries to replicate the effect using conversational speech data from a different variety of English (i.e. New Zealand English), using the QuakeBox Corpus (Walsh et al. 2013). In the presence of word-form frequency as a predictor, stems of plurals were not found to be significantly longer than the corresponding strings of comparable non-complex words. The analysis revealed, however, a frequency-induced gradient paradigm uniformity effect: plural stems become shorter with increasing frequency of the bare stem.

Keywords

paradigm uniformity, acoustic duration, morphology, articulation, phonetic reduction

1. Introduction[1]

Recent work on the acoustic properties of complex words has found that morphological information may influence the phonetic properties of words, for example acoustic duration. For English, a number of studies have provided evidence for such effects on stems and affixes (e.g. Plag, Homann & Kunter 2017; Seyfarth et al. 2017; Tomaschek et al. 2019; Plag et al. 2020; Ben Hedia & Plag 2017; Ben Hedia 2019; Hay 2007; Plag & Ben Hedia 2018; Lee-Kim, Davidson & Hwang 2013; Mackenzie et al. 2018; Bell, Ben Hedia & Plag 2019).

It is currently not quite clear how such effects come about. One explanation, put forward by Seyfarth et al. (2017) for effects on the duration of inflectional stems, is paradigm uniformity. These authors found that stems of words ending in [s, z] have longer durations if these are inflected words, whereas the corresponding strings of segments in mono-morphemic words ending in [s, z], henceforth 'pseudo-stems', have shorter durations. They argue that these differences in stem duration are due to a paradigm uniformity effect, in which the stem of a morphologically complex word like *days* is influenced by its morphologically simple paradigm member *day*. In a nutshell, stems like *day* have an open syllable and are at the edge of a prosodic boundary (the prosodic word), and therefore show a lengthening effect. This longer duration influences the articulation of the complex form *days*. Supposedly, the effect comes about through the co-activation of the articulatory plan of the morphologically related stem *day* when the speaker tries to articulate *days*. Crucially, there is no morphologically induced co-activation of *day* when the speaker tries to articulate the mono-morphemic form *daze*, hence lengthening does not occur with *daze*.

However, Seyfarth et al. (2017)'s experimental results only partly confirm the alleged paradigm uniformity effect. There is an effect for final /s/ and /z/ such that pseudo-stems are shorter than suffixed stems, but there was no such effect for words ending in final /t/ and /d/, involving the past tense suffix. Furthermore, Seyfarth et al. (2017) tested a further prediction emerging from paradigm uniformity, based on work by Winter & Roettger (2011) and Roettger et al. (2014): a stronger representation of the stem (as gauged by its frequency) should lead to an even longer duration of the suffixed stem. This turned out to be not the case. There was no relation between the absolute or relative frequency of the bare stem and the duration of the suffixed stem.

To address some of the problems raised by Seyfarth et al. (2017)'s results, and to test paradigm uniformity in casual speech, the present study investigates paradigm uniformity using

[1] The data set used for this study and the script for the statistical analysis are available at https://osf.io/p9rv6/?view_only=d100d29a96f24b78852dd231c7700ef5

data from a speech corpus containing natural conversations, the QuakeBox Corpus (Walsh et al. 2013), which was recorded in New Zealand. Using natural speech presents its own challenges, but is motivated by the need to replicate effects that have been observed under laboratory conditions under the conditions of everyday language use (see Tucker & Ernestus (2016) for discussion).

In our data set we find that stems of plural words ending in [z] are not significantly longer than pseudo-stems of mono-morphemic words ending in [z], i.e. we cannot replicate the paradigm uniformity effect found by Seyfarth et al. (2017).

Like Seyfarth et al. (2017), we do not find evidence that more frequent bare stems exert a stronger lengthening effect on plural stems. Interestingly, we find the opposite being the case. In the conversational data, increasing bare stem frequency goes together with decreasing plural stem duration. We argue that this effect is in accordance with paradigm uniformity. Independently of any paradigmatic effect, higher frequency of the bare stem quite expectedly leads to a shorter articulation of the bare stem. This shorter articulation influences the duration of the plural stem via the articulatory plan as explained above, leading to shorter plural stems for words with increasing bare stem frequency. It is thus not the strength of the lexical representation that would enhance the paradigmatic influence in the direction surmised by Seyfarth et al. (2017). It is the articulatory plan of the phonetically reduced high frequency bare stem that makes the plural stem also shorter.

This paper is structured as follows; in section 2 we will look at the influences of morphology on speech production, and at paradigm uniformity (and similar) effects. In section 3 we introduce the corpus, explain the variables used, and introduce the statistical analysis. In section 4 we present the results of the statistical analysis, and in section 5 we summarize our results and discuss their theoretical implications.

2. Morphology and speech production

Human speech is a vastly complex process which is influenced by many different non-linguistic and linguistic factors. Among these are for example the frequency of a word (e.g. Jurafsky et al. 2001; Pluymaekers, Ernestus & Baayen 2005; Pluymaekers, Ernestus & Baayen 2005; Gahl 2008; Bell et al. 2009; Lohmann 2018), the prosody of a sentence or word and the position of a word within an utterance (e.g. Wightman et al. 1992; Fougeron & Keating 1997; Tabain 2003), the age, gender or social or regional origin of a speaker (e.g. Labov 1972; Byrd 1994), the predictability of a particular word within a given context (Bell et al. 2009), or the morphological structure of a word (e.g. Hay 2003; Kemps et al. 2005; Lee-Kim, Davidson & Hwang 2013; Blazej & Cohen-Goldberg 2015; Ben Hedia & Plag 2017; Plag, Homann &

Kunter 2017; Plag, Engemann & Kunter 2018a; Plag, Engemann & Kunter 2018b; Pluymaekers et al. 2010; Seyfarth et al. 2017; Tomaschek et al. 2019; Plag et al. 2020a; Engemann & Plag 2020; Engemann, Plag & Zimmermann 2019; Zee 2019; Schmitz, Plag & Baer-Henney 2020; Plag et al. 2020b).

A number of recent studies constitute a growing body of evidence on the interaction between morphology and phonetics, more specifically, how morphological structure may affect the acoustic properties of stems in complex words. In a study on Dutch singular and plural nouns, Kemps et al. (2005) found that inflected and uninflected forms have different acoustic, durational characteristics. Stems in plural words were on average about 90 milliseconds shorter than singular words and listeners were sensitive to these durational differences between singular forms and the stems of plural forms. This means that plurals are not just singulars with an additional suffix, but that the acoustic realization of plurals may be influenced by their morphological structure.

Cohen (2014) investigated words with morphemic final [s] and [z] in English, and found that the duration of these sounds and the stems can vary dependent on morphological properties such as paradigmatic probability. Higher paradigmatic probability causes longer suffixes, as well as shorter stem durations.

Caselli, Caselli & Cohen-Goldberg (2016) approached the topic of durational differences between inflected and mono-morphemic words from the perspective of phonological neighborhood density, arguing that word production is influenced not only by phonological neighborhood density itself, but also by inflected neighborhood density, a measurement that they define as the number of inflected words that differ from a target word by one phoneme. The authors also found that both word-form frequency and stem frequency influence acoustic duration. Increasing word-form frequency and increasing bare stem frequency go together with shorter word durations.

3. Paradigm uniformity

In morphology, a paradigm is a set of morphologically related forms. In derivational morphology a paradigm may consist of all forms with a specific affix (also known as ‘morphological category’), or of the derived words that share a given root (also known as ‘morphological family’). In inflection, a paradigm contains all inflected word-forms of a given lexeme (or ‘lemma’). For example, the inflected forms *free*, *frees*, *freed*, *freeing* constitute the morphological paradigm of the verb lexeme FREE.

A paradigm uniformity effect arises when a morphologically complex form is influenced by other members of its paradigm.[2] In the theoretical-linguistic literature, derivational paradigm uniformity has been suggested to account for quite a number of cases where paradigms show unexpected variability.

Consider variable stress patterns in English complex words. For example, the word *demonstrable* may be produced with varying stress: *démonstrable* or *demónstrable*. In a paradigm perspective this variability may arise through the competition of at least two morphologically related forms: *démonstrate* and *demónstrative* (Bauer, Lieber & Plag 2015).

Paradigm uniformity effects have been discussed for a number of phenomena also in other languages (e.g. Greek: Gafos & Ralli (2002), Korean: Park (2006); Kenstowicz & Sohn (2008), French: Bonami et al. (2019), Hebrew: Laks, Cohen & Azulay-Amar (2016), Hungarian: Rebrus & Törkenczy (2005), Luwanga: Green (2009), Russian: Bethin (2012)), mostly at the level of phonology. However, parts of the discussion have focused on non-contrastive effects, raising the question of phonetic versus phonological effects in morphology. A phenomenon that has attracted particular attention is incomplete neutralization, e.g. in English r-flapping (Steriade 2000; Riehl 2003; Eddington 2006; Braver 2014) or final devoicing in Dutch and German (e.g. Ernestus & Baayen 2006; Winter & Roettger 2011; Roettger 2014).

In German, words such as *Rad* 'wheel' and *Rat* 'council' are considered homophonous in pronunciation since the underlying voicing contrast between the two forms (as represented also in the spelling) is neutralized due to German final obstruent devoicing. However, studies (e.g. Winter & Roettger 2011; Roettger et al. 2014)) have found that there are fine phonetic differences between these homophonous pairs, which may be due to the influence of the morphologically related forms, such as *Räder* 'wheel-PLURAL', or *Rades* 'wheel-GENITIVE'. When a speaker produces *Rad*, morphologically related forms are also activated, and this affects the articulation of the word in question. In the case of *Rad*, this results in an incomplete devoicing of the final obstruent.

In the case of English verbal paradigms, different forms may also influence each other, for example *frees* may be phonetically influenced by the related form *free*. In this particular case, this may happen as follows: Since *free* has an open syllable, and this syllable is at the end of a prosodic domain (i.e. the prosodic word) it is pronounced with a phonetically rather long vowel, i.e. with a longer articulatory gesture for the vowel and perhaps also its preceding material. Due to co-activation in lexical processing, this stored articulatory gesture influences other paradigm

[2] In the linguistic literature, the phenomenon is also known as 'stem selection' (Raffelsiefen 2004: 95), 'multiple correspondence' (e.g. Burzio 1998), or the 'split-base' effect (Steriade 2000).

members when these are pronounced, causing the stems of *frees*, *freed* or *freeing* to be pronounced with a relatively long duration. This effect is especially visible in homophonous word pairs, as investigated by Seyfarth et al. (2017). These authors found the stem of /z/- or /s/-suffixed stems like *frees* to be longer than the equivalent phonetic material in corresponding mono-morphemic words (such as *freeze*), which are not influenced by the bare stem *free* because of the lack of a morphological relation to *free.* Seyfarth et al. (2017) propose that paradigm uniformity is the cause of the durational differences that can be observed in their results. It should be noted, however, that, for unclear reasons, the effect was absent from the /d/- and /t/-suffixed stems.

The question of subphonemic effects of paradigm uniformity is part of a larger discussion of the important question of how lexically related forms may influence each other in speech production (Ernestus & Baayen 2006; Goldrick & Blumstein 2006; Roettger et al. 2014; Dell 1986; Goldrick 2014; McMillan, Corley & Lickley 2009; Peterson & Savoy 1998; Rapp & Goldrick 2000; Winter & Roettger 2011). This is still an open question, and a robust effect of paradigm uniformity would be an important finding to feed into this discussion.

It is, however, not quite clear, how robust such effects really are. Seyfarth et al. (2017) themselves do not find the effect for /d/- and /t/-suffixed stems. Frazier (2006) finds durational differences between homophonous past tense and mono-morphemic forms (such as *band* vs. *banned*) in the direction predicted by categorical paradigm uniformity. However, the statistical analysis is inadequately documented and the results inconclusive.[3] Seyfarth, Vander Klok & Garellek (2019) looked at Javanese verbal paradigms and did not find the expected categorical paradigm uniformity effects concerning the phonetic parameters investigated (nasal resonance and closure duration).

It is thus necessary to replicate the effect from Seyfarth et al. (2017) with other data sets. Furthermore, the effect should not be restricted to laboratory speech, but should also be found in spontaneous speech production in natural conversations. It has been argued, e.g. by Tucker & Ernestus (2016) that research on speech production needs to shift its focus to spontaneous speech to be able to draw valid conclusions about language processing. For the present paper, we have chosen the Quakebox Corpus for the investigation of paradigm uniformity effects.

Seyfarth et al. (2017) investigate two possible effects of paradigm uniformity, which we will refer to in this article as 'categorical' paradigm uniformity and 'gradient' paradigm uniformity. The categorical effect is an effect that holds across the board: categorically, stems of words that

[3] The paper gives some means and some results of z tests. There is no state-of-the-art analysis using statistical methods that take into account random effects or important co-variates like speech rate.

end in word-final [s, z] are longer when these words are inflected, such as in the case of *frees*, whereas morphologically simple words, like the homophonous *freeze*, have shorter pseudo-stems.

Following Winter & Roettger (2011) and Roettger et al. (2014), Seyfarth et al. (2017) hypothesize that the strength of the paradigm uniformity effect depends on frequency. Based on the reasoning of what causes the acoustic difference in question, it is predicted that the influence of the bare form on a suffixed form becomes stronger with increasing strength of representation of the bare stem. One correlate of this strength is frequency, such that more frequent bare stems show a stronger lengthening effect on the suffixed form. The frequency of the bare stem can be measured in absolute or in relative terms. Relative frequency is the ratio of the frequency of the suffixed form and the bare stem. If the bare stem is relatively less frequent (as for example *shoe* as against *shoes*) relative frequency is high. Higher relative frequency would mean that the influence of the bare stem is weaker (cf. Zuraw & Peperkamp 2015), hence the stem of the inflected word would be shorter.

While Seyfarth et al. (2017) found robust evidence for a categorical paradigm uniformity effect with words ending in [s, z], they found no evidence for a gradient paradigm uniformity effect. Neither of the two frequencies significantly correlated with acoustic duration of the stem. Seyfarth et al. (2017) do not discuss their null result concerning gradient paradigm uniformity. They only remark that their results "should be interpreted with caution, in particular because the stimuli were not selected to include a broad range of either frequency measures" (p. 9).

It is not clear, however, whether Seyfarth et al. (2017)'s hypothesis concerning a gradient paradigm uniformity is conceptually on the right track. It seems to make sense that a stronger representation of the bare stem may exert a greater influence on morphologically related forms. However, the direction of the effect may be opposite to the one expected by Seyfarth and colleagues due to effects of phonetic reduction that go hand in hand with rising lexical frequencies. It is well known that more frequent words tend to have shorter realizations. This has been demonstrated for lemma frequency (e.g. Jurafsky et al. 2001; Bell et al. 2009; Gahl 2008; Lohmann 2018) and for word-form frequency (Caselli, Caselli & Cohen-Goldberg 2016; Lõo et al. 2018). For instance, Caselli, Caselli & Cohen-Goldberg (2016) found that the stem frequency of words inflected with *-ed* and *-ing* negatively correlates with the duration of these words in speech.

Based on these findings on phonetic reduction, an alternative hypothesis concerning gradient paradigm uniformity suggests itself. The more frequent a bare stem, the shorter its duration. This relatively shorter duration, or rather the stored concomitant articulatory plans, should

influence the plural stems in such a way that plural words with a more frequent bare stem should also have shorter stems than plural words based on less frequent bare stems. The frequency of the bare stem would therefore not only have an effect on duration of the bare stem itself, but also indirectly on the duration of this stem when part of a plural form.

In addition, we can expect an effect of plural word-form frequency itself. The more frequent the plural word-form, the shorter the duration of this word-form and thus of the stem it contains.

But what about relative frequency? According to our alternative hypothesis, high values of plural word-form frequency go together with shorter plural stems, and high values of bare stem frequency also go together with shorter plural stems. Increasing values of relative frequency indicate either higher plural frequency, and hence shorter durations, or lower bare stem frequency, and hence longer durations. If we control for plural frequency, we should expect longer durations with increasing relative frequency.

To summarize, our study sets out to investigate both categorical and gradient paradigm uniformity effects using conversational speech, concentrating on plural as the morphological category of interest. In particular, we test the hypotheses given in (1)-(3):

(1) H1: **Categorical paradigm uniformity**

Stems of plural words ending in the suffix [z] have longer durations than the pseudo-stems of mono-morphemic words ending in [z].

(2) H2: **Gradient paradigm uniformity due to activation strength** (Seyfarth et al. 2017)

a.) The higher the absolute frequency of the bare stem, the longer the duration of the plural stem.

b.) The higher the relative frequency, the shorter the stem of the inflected word.

(3) H3: **Gradient paradigm uniformity due to phonetic reduction**

a.) The higher the absolute frequency of the plural word-form, the shorter its duration, and the shorter the duration also of its stem (general reduction effect).

b.) The higher the absolute frequency of the bare stem, the shorter the duration of the plural stem.

c.) The higher the relative frequency (and keeping plural word-form frequency constant), the longer the duration of the plural stem.

4. Methodology

4.1 Data set

For the present study we used the QuakeBox Corpus (Walsh et al. 2013). This corpus was recorded in Christchurch, New Zealand, after the earthquakes in 2010 and 2011, which destroyed large parts of Christchurch and surrounding towns. The corpus consists of monologues in which speakers share their experiences during and after the earthquakes, most of which are emotional or even traumatic. Most speakers in this corpus are between 36 and 65 years old, and 65% of speakers are female, while 35% are male.

For our analysis, we made use of a dataset that was extracted from the QuakeBox Corpus (Walsh et al. 2013) by Julia Zimmermann in order to study the durations of different types of word-final S (Zimmermann 2016). She first extracted all words ending (phonemically) in word final /s/ or /z/ that were not followed by an S-initial word (as these initial S's tend to merge with word-final S), and that were produced by speakers identifying as ethnically New Zealand European. Subsequently, this dataset was cleaned by excluding brand names, place names and clitics that do not represent *has* or *is*. Furthermore, words from word classes other than nouns, verbs and pronouns were eliminated, as well as function words such as *has*, *is*, *was*, etc. To avoid over-representation of certain lexemes, only up to 25 tokens of each combination of base and type of S were randomly sampled. Finally, Zimmermann excluded items with a final S that had a duration longer than 250 milliseconds, items with a speech rate faster than 15 syllables per second, and items for which the analysis of center of gravity showed obviously false measurements (see Zimmermann 2016 for details). The resulting dataset contained 7073 tokens.

For the purpose of investigating paradigm uniformity, we further reduced this dataset to achieve greater homogeneity concerning the words to be investigated. We only included monosyllabic words with a vowel preceding the word final [s] or [z] because of the distribution of these consonants. Having only vowel-final stems had the consequence that only [z]-final words occur in our data set, since stem-final vowels, being obligatorily voiced, trigger the voiced allomorph of the plural, i.e. [z]. We therefore also included only mono-morphemic words that ended in [z]. We only included plural and mono-morphemic words ending in the same vowel + [z] structure.

The inspection of the distribution of the stem durations showed the presence of some outliers. Based on visual inspection of the distribution, we removed observations that had stem durations longer than 607 ms, or shorter than 135 ms. This eliminated 36 observations.

The resulting data set had 487 tokens and 74 types, out of which 34 types with 163 tokens are mono-morphemic words, and 40 types and 324 tokens are plural words. In order to be able

to take into account correlated errors for word types in a mixed effects regression analysis we only used words that were attested at least 3 times in the data set. This reduced the data set to 431 tokens representing 38 types. In this data set, there are 136 mono-morphemic word tokens (18 types), and 295 tokens (20 types) of plural words.

For lists of types and number of tokens, see appendix. This data set ('data set 1') was used to test categorical paradigm uniformity. We created a statistical model using this dataset in which the log-transformed duration of the (pseudo-)stem is the response variable and the morphological make-up (mono-morphemic as against plural) and log-transformed word-form frequency are the predictor variables of interest.

For the investigation of gradient paradigm uniformity, we eliminated the mono-morphemic words from the data set, since we are only interested in the relationship between plurals and their stems. This data set ('data set 2') had 20 types and 295 tokens. With this dataset, we investigated how the frequency of the bare stem, the frequency of the plural word-form, and the relative frequency of bare stem and plural form affect the duration of the suffixed stem.

For the statistical analysis we used multiple linear mixed effects regression with WORD as random intercept to control for by-word variation. It was not possible to include speaker as a random effect because the data set with 431 tokens originates from the speech of 196 speakers, the majority of which contributed only one token. In order to control for a very important speaker-dependent variable, articulation rate, we included local speech rate as a co-variate (see below).

4.2 Variables

4.2.1 Response Variable: Duration of stem

The response variable STEMDURATION is the log-transformed duration of all phonemes minus the word-final [z]. In a morphologically complex word such as *days* [deɪz], this corresponds to the duration of the phones [deɪ], i.e. the stem. In a mono-morphemic word such as *daze* [deɪz], this measurement also corresponds to the duration of the phones [deɪ], i.e. the pseudo-stem. Relevant acoustic measures such as duration and voicing were extracted automatically from the corpus with the help of LaBB CAT (Fromont & Hay 2012) and a script for the acoustic analysis software Praat (Boersma & Weenink 2015). Durations were measured based on the corpus' automatically aligned phonetic transcriptions.

4.2.2 Variables of interest

MORPHEMETYPE represents the morphological make-up of the word under investigation. The value of this variable may either be `S` (for mono-morphemic words) or `PL` (for plural

words). Like all other categorical variables in our data set, MORPHEMETYPE was coded using the default coding available in R, that is dummy coding.

WORDFORMFREQ is the word-form frequency of the word under investigation, such as the frequency of *days* as it appears in its plural form, or the frequency of the form *daze*. These frequencies were extracted from the BNC and log-transformed.

STEMFREQ is the frequency of the stem of the complex word under investigation, for instance the frequency of *day* (for the complex word *days*). These frequencies were also extracted from the BNC and log-transformed.

RELATIVEFREQ is calculated, as in Seyfarth et al. (2017), by dividing the word-form frequency by the bare stem frequency. The resulting variable was then log-transformed. Note that this variable is only available for the morphologically complex words.

4.2.3 Control variables

Due to the intricacies of human speech, a large number of additional variables may affect the analysis of speech data. To control for some of these variables, we included several fixed effects in our statistical models, their selection following similar studies (e.g. Gahl (2008); Plag, Homann & Kunter (2017)).

NUMPHON. To control for differences in the phonological makeup we included the number of phonemes of the word in question. This variable was created by counting the number of phonemes that were in the phonemic transcription provided by the corpus.

EXPSTEMDUR. Individual segments differ in duration, and it is therefore desirable to control for these durational differences (for instance the difference between a lax and a tense vowel). To address this concern we used a procedure analogous to that used in Gahl, Yao & Johnson (2012) and Caselli, Caselli & Cohen-Goldberg (2016). We first calculated the average duration of each segment across the whole QuakeBox corpus. We then used these average segment durations to calculate the expected duration of a given stem (i.e. of the word minus the duration of the final /z/) by adding up the average durations of the respective segments. These baseline durations were log-transformed.

SPEECHRATE is provided as meta-data by the corpus and was adopted as is. It was calculated by the corpus compilers in syllables per second for each utterance, i.e. between pauses or turn boundaries. This variable thus controls for speaker-specific and utterance-specific articulation rate. This variable was also log-transformed.

LBIGRAMPROB and RBIGRAMPROB. The probability of a word in its immediate context is another influential factor for duration; studies have shown that the preceding or upcoming context of a word can affect its acoustic duration. (e.g. Bell et al. 2003; Pluymaekers, Ernestus

& Baayen 2005a; Torreira & Ernestus 2009). We used bigram probabilities estimated on the basis of the BNC. To be able to include bigrams with a frequency of 0 we added 1 to all bigram frequencies. The left and right bigram probabilities were log-transformed and labeled LBIGRAMPROB and RBIGRAMPROB.

NEIGHBORDENSITY and NEIGHBORFREQUENCY. Neighborhood densities and neighborhood frequencies can influence phonetic duration (e.g. Gahl, Yao & Johnson 2012). Both neighbourhood measures were extracted from the CLEARPOND database (Marian 2012). NEIGHBOURDENSITY refers to the number of words differing in one segment from the item in question, while NEIGHBOURFREQUENCY is the mean frequency (per million) of these neighbouring words.

POSITION contains the location of the word within the sentence. Due to phrase-final lengthening, segments at the end of prosodic constituents tend to be pronounced longer (Klatt 1976; Byrd, Krivokapic & Lee 2006). The corpus provides information about the position of a word in an utterance, with the levels (`middle`, `pause`, `nearpause`, `falsestart`, `hesitation`, `nearfinal`, `final`). The vast majority of words in the dataset are in the `middle` position where we expect shorter durations than before pauses. The values `nearpause` and `nearfinal` refer to items in which the following word is final or before a pause. The difference between `pause` and `final` is not quite clear from the corpus description and the two values seem to be employed by the transcribers in such a way that they seem interchangeable. To address potential issues with the number and operationalization of the seven values as they come with the corpus annotation, we recoded this variable into a binary variable, with the values `middle` and `non-middle`. The value `middle` remained as is, and the new value `non-middle` conflated all other values. Models using this new position variable instead of the original variable were nearly identical to models using the original variable. It was therefore decided to use the original variable with its more fine-grained values.

Part-of-speech (PARTOFSPEECH) contains information about the word class, such as noun or verb, and was included early in the modelling process for Categorical Paradigm Uniformity, but was found to be not significant.[4] In the analysis of Gradient Paradigm Uniformity this variable is superfluous because those data sets only contain plural words, i.e. nouns.

[4] Note that in Seyfarth's study the mono-morphemic words also came from different parts-of-speech (noun, verb, adjective, adverb). The authors did not control for this potential source of variation in their statistical modeling. A related problem is the fact that many words, in both Seyfarth et al.'s and our study, are ambiguous. For instance, Seyfarth et al. have a third person singular verb form *paws* among their critical stimuli that is homophonous with the plural of the noun *paw* (which is not tested in the study). Similarly, their plural item *brews* is also a third person singular form of the verb *brew* (not included in the study), and the corresponding form *bruise* is ambiguous between a verb and a noun. The problem is ignored in Seyfarth et al's study, as well as in ours.

AGEGROUP contains the age of the speaker as a numerical value ranging from 1 to 7. These values correspond to the age spans of 18-25, 26-35, 36-45, 46-55, 56-65, 66-75, 76-85 and 85+. Older speakers are expected to have lower speech rates, hence longer durations (Ramig Lorraine A. & Ringel Robert L. 1983; Skoog Waller, Eriksson & Sörqvist 2015).

VOICERATIO contains the number of frames of the word final [z] that show vocal fold vibration divided by the total number of frames of the word final [z]. This accounts for phonetic differences in the duration of voicing. Voiced [z] is shorter than unvoiced [s] (e.g. Klatt 1976), which might also affect the duration of the stem in some way.

4.2.4 Summary of variable distributions

In Table 1 and Table 2 we present summaries and distributions of the above described variables.

Table 1: Summary of the dependent variables, predictors and covariates for data set 1

Dependent variable	N	Mean	St. Dev.	Min	Max
STEMDURATION	431	-1.246	0.310	-1.966	0.511
Numerical variables					
WORDFORMFREQ	431	8.521	1.180	4.700	10.373
SPEECHRATE	431	1.393	0.162	0.806	1.859
VOICERATIO	431	0.390	0.245	0.000	1.000
NUMPHON	431	3.181	0.487	2	4
EXPSTEMDUR	431	-1.613	0.260	-2.549	0.984
LBIGRAMPROB	431	4.253	2.483	0.000	8.567
RBIGRAMPROB	431	3.942	2.493	0.000	8.787
AGEGROUP	422	4.123	1.562	1.000	7.000
NEIGHBORDENSITY	427	26.166	10.716	8.000	50.000
NEIGHBORFREQUENCY	427	127.073	318.906	4.796	33.183
Categorical variables	**N**	**Levels**			
WORD	431	38 levels			
SPEAKER	431	196 levels			
POSITION	431	pause: 54	falsestart: 3	final: 19	hesitation: 7
		middle: 325	nearfinal: 11	nearpause: 21	
PART OF SPEECH	389	AJ0: 4	NN1: 71	NN2: 292	VVB: 22
MORPHEMETYPE	431	mono-morphemic: 136		plural: 295	

Table 2: Summary of the dependent variables and covariates for dataset 2

Dependent variable	N	Mean	St. Dev.	Min	Max
StemDuration	295	1.218	0.301	1.966	0.511
Numerical variables					
wordFormFreq	295	8.535	1.087	5.517	10.373
stemFreq	295	9.429	1.416	6.455	11.936
relativeFreq	295	0.894	1.210	6.418	1.134
SpeechRate	295	1.390	0.160	0.842	1.859
voiceRatio	295	0.368	0.230	0.000	1.000
numPhon	295	3.176	0.505	2	4
ExpStemDur	295	1.627	0.216	1.987	0.984
LBigramProb	295	4.461	2.468	0.000	8.567
RBigramProb	295	3.793	2.263	0.000	8.121
ageGroup	287	4.118	1.549	1.000	7.000
neighborDensity	295	25.580	8.997	9	44
neighborFrequency	295	154.585	377.098	11.066	1,687.658

Categorical variables	**N**	**Levels**			
word	295	20 levels			
speaker	295	164 levels			
position	295	pause: 30	falsestart: 2	final: 11	hesitation: 3
		middle: 227	nearfinal: 8	nearpause: 14	

4.3 Modeling procedure

We analyzed our data using multiple mixed effects linear regression in the statistics language and program R (R version 3.6.1 (2019-07-05), R Core Team 2019). Mixed-effects regression brings the variation of random effects, such as subject or item, under statistical control and can deal with unbalanced data sets. The latter property is especially welcome in corpus analyses since usually not all combinations of all values of the different predictors are represented with equal frequency in samples drawn from a corpus. In addition, we also used random forests to address concerns of collinearity (e.g. Tomaschek, Hendrix & Baayen (2018)).

We first fitted a model to predict stem duration on the basis of MorphemeType. This allowed us to test for a categorical paradigm uniformity effect. We then fitted models to test for a gradient paradigm uniformity effect, testing the effects of StemFreq, WordFormFreq and RelativeFreq.

The models were fitted according to the following strategy: In the initial model, we included all control variables alongside the variable of interest as fixed effects. In addition, we used the word type (Word) as the random effect. In the model testing the effect of MorphemeType (i.e.

the model using data set 1), we also tested interactions of all other variables with MORPHEMETYPE. The models then went through a standardly used step-by-step elimination process (e.g. Baayen 2008), in which we reduced the number of predictor variables based on their predictive power in the models. A variable had to pass three tests to be included in a model. First, it had to yield a t-value greater than 2 (or less than -2). Second, the Akaike information criterion (AIC) of the model including the variable had to be lower than the AIC of the model without it. Third, a likelihood ratio test comparing the model including the factor to a model without it had to yield a *p*-value lower than 0.05, thus showing that the inclusion of the factor did significantly improve the fit of the model. A variable under consideration was only retained in the model if it passed all three tests.[5] Variable elimination proceeded in such a way that in each new model the variable with the highest *p*-value was tested first. To address collinearity issues, various measures were taken. These will be explained as we go along. The residuals of the final regression models showed a normal distribution in both tails of the distribution.

5. Results

5.1 Categorical paradigm uniformity

A comparison of the durations of pseudo-stems of mono-morphemic words and stems of plural words shows a difference in the medians of 30 ms, with plural stems being longer (plural: 290 ms, mono-morphemic: 260 ms). To control for potentially confounding variables, a model was fitted according to the above described procedure using dataset 1 and all predictors described above (apart from bare stem frequency, for which the mono-morphemic words do not have values). The difference in stem duration between two morphological categories are no longer significant in the presence of the other predictors. In the model that contained the remaining significant predictors and only MORPHEMETYPE as predictors, MORPHEMETYPE reached a *p*-value of 0.25 (estimate: -0.069, *t*=-1.172*, df*=28.59*, strd. error*=0.059). Table 3 gives the output of the model that contains only the remaining significant predictors, and Figure 1 plots the partial effects of this model. The fixed effects explain 23 percent of the variance, the entire model 46 percent (based on pseudo-R-squared for Generalized Mixed-Effect models, using the function `r.squaredGLMM()` from the `MuMIn` package, Barton (2009)). None of the interactions of the covariates with MORPHEMETYPE were significant.

[5] The reader should note that a variable may pass all three tests and still may not quite have a significant p-value in the regression model output. This happened with two of our models (see below).

Table 3: Fixed-effects coefficients and p-values in the final model testing the categorical paradigm uniformity effect. (Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05)

Random effects						
Group	Name		Variance		Std. Dev.	
WORD	(Intercept)		0.023		0.15	
Residual			0.056		0.24	
Fixed effects						
	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.25	0.27	56.42	0.95	0.35	
WORDFORMFREQ	-0.06	0.02	35.63	-2.60	0.01	*
EXPWORDDUR	0.33	0.09	37.56	3.52	0.00	**
POSITIONfalsestart	-0.13	0.15	394.38	-0.89	0.38	
POSITIONfinal	0.00	0.07	388.84	-0.04	0.97	
POSITIONhesitation	-0.05	0.10	390.48	-0.47	0.64	
POSITIONmiddle	-0.17	0.04	394.70	-4.08	0.00	***
POSITIONnearfinal	-0.10	0.08	385.14	-1.23	0.22	
POSITIONnearpause	-0.13	0.07	395.23	-1.78	0.08	.
SPEECHRATE	-0.26	0.08	400.47	-3.33	0.00	**
VOICERATIO	-0.21	0.05	397.66	-3.90	0.00	***
AGEGROUP	0.02	0.01	394.35	2.12	0.03	*

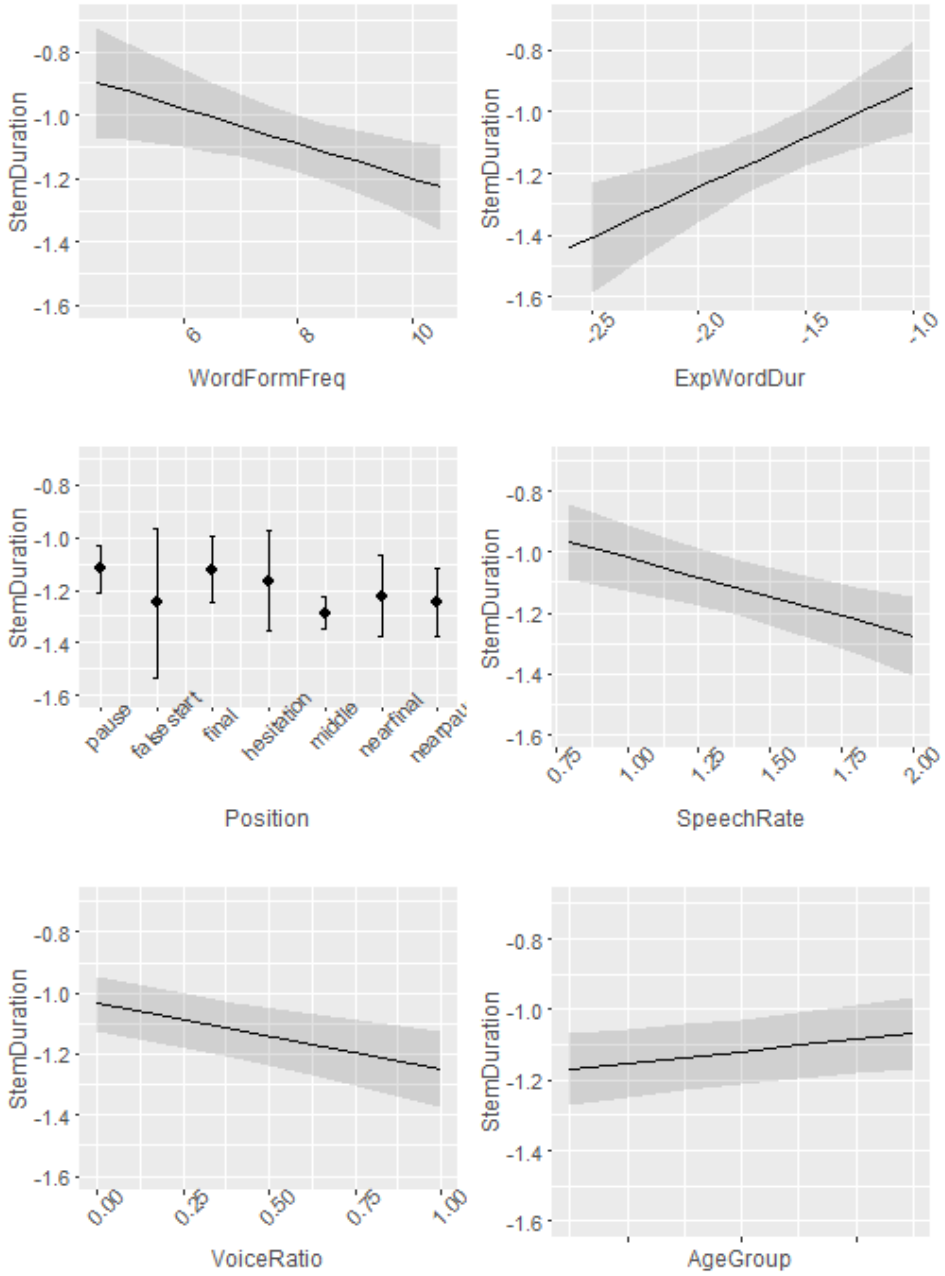

Figure 1: Partial effects of the final model testing paradigm uniformity

The effects of the covariates pattern as expected. The estimate for the variable WORDFORMFREQ indicates that the higher the frequency of a word-form within the BNC corpus, the shorter the duration of the stem or pseudo-stem. This is an effect that has been well established in the literature, both on the lemma level and on the word-form level (e.g. Jurafsky et al. 2001; Pluymaekers, Ernestus & Baayen 2005b; Gahl 2008; Bell et al. 2009; Lohmann 2017; Lõo et al. 2018; Plag et al. 2020a). The more often a word is used within a language, the shorter its duration, whereas words that are used more infrequently tend to have longer durations. Another effect that is well established in the literature is phrase-final lengthening, which can also be found in our analysis when inspecting the coefficient of POSITION (note that

the baseline condition for this variable in Table 3 is `pause`). The variable SPEECHRATE also behaves in an unsurprising way: the faster the speech rate of the speaker, the shorter the duration of the stem. The effect of EXPWORDDUR, i.e. baseline duration, is also unsurprising. Words with longer phonemes have longer stem durations. The higher the value of VOICERATIO, the more voicing is present in the final /z/, the shorter the /z/ and the shorter the duration of the stem. In other words, durational reduction affects whole words, including the final fricative. This effect is unsurprising and has been observed also in other studies of words involving final /z/ or /s/, for example Plag, Homann & Kunter (2017); Plag et al. (2020). Finally, the variable AGEGROUP shows that the older the speaker, the longer the duration of the stem, indicating that older speakers tend to speak slower in general than younger speakers. Again, this is an expected effect.

5.2 Gradient paradigm uniformity

Seyfarth et al. (2017) tested for gradient paradigm uniformity by assuming strength of activation as the responsible mechanism that drives the magnitude and direction of the effect. The higher the strength of activation, the stronger the lengthening of the morphologically related form. They devised two different models, each with only one type of frequency as the predictor of interest. That is, in one model bare stem frequency was included as the sole predictor of interest, in the other model relative frequency was included as the sole predictor of interest. Plural word-form frequency was not included in their models, although word-form frequency has been shown to influence word duration, and thus, presumably, stem duration.

Since we are testing Seyfarth's hypothesis as well as our alternative hypotheses, all three frequencies are of interest for the present study. Given that the three frequencies may correlate with each other, this means that we should first have a look at potential collinearity.

Table 4 presents a correlation matrix for the three variables.

Table 4: Correlation matrix for lexical frequency measures in data set 2 (rho-values, p-values are given in parentheses, Spearman test)

	RELATIVEFREQ	STEMFREQ
WORDFORMFREQ	0.14 (0.02)	0.57 (0.00)
RELATIVEFREQ		-0.67 (0.00)

Including all three frequencies is ill-advised, as STEMFREQ is strongly correlated with both WORDFORMFREQ and RELATIVEFREQ. Using the condition number as a measure of collinearity danger for these three variables (with the `collin.fnc()` from the `languageR` package (Baayen & Shafaei-Bajestan 2019)), we get an extremely high figure of 19,701,789,290. Values above

30 are considered to indicate harmful collinearity (e.g. Tomaschek, Hendrix & Baayen (2018)). Pairwise calculation of condition numbers leads to acceptable values ranging from 18.3 to 20.2.

We also used variance inflation factors (VIFs) (using the `vif()` function of the car package (Fox & Weisberg 2011)). When we tried to apply `vif()` to a linear model with all three frequencies, the linear model can only estimate two of the three coefficients. Collinearity is a likely reason why including all three variables as predictors into the regression models leads to rank deficiency, such that only two of three coefficients can be estimated. If we include only two of the three frequency variables (in addition to all covariates) in the linear model, the resulting VIFs for the frequency variables are between 2.1 and 3.6. As a rule of thumb, values below 5 are considered acceptable (e.g. Tomaschek, Hendrix & Baayen 2018).

There are different strategies available to address collinearity issues (see, e.g., Tomaschek, Hendrix & Baayen (2018)). In our case, with only three variables at issue, where one of the three is even calculated on the basis of the other two, and where including all three at the same time is impossible, three strategies suggest themselves, as shown in the subsequent sections.

5.2.1 Testing the effects of different frequency measures on stem duration

To test Seyfarth's hypotheses, one possibility is to mirror their procedure and devise models with only either relative frequency or bare stem frequency as predictor of interest (i.e. without word-form frequency as a covariate). These models were fitted initially including all covariates, and according to the simplification procedure described above. In the model with relative frequency, this predictor is not significant (initial model: t=1.14 p=0.27). In contrast, in the model with bare stem frequency this predictor remains significant in the final model (*coefficient*=-0.075, *std. error*=0.027, t=-2.76, p=0.015). The negative coefficient of bare stem frequency shows that with rising bare stem frequency the duration of the stem in plural forms becomes shorter. The direction of this effect is in the opposite direction from Seyfarth's hypothesis (H2a) and confirms the effect expected based on the considerations underlying the alternative hypotheses (H3b).

These models (like Seyfarth et al.'s) are, however, flawed by not including word-form frequency as a covariate. Given the acceptable condition numbers and VIFs for models containing two of the three measures, it seems safe (even if not ideal) to include word-form frequency in the two models. This additional variable turns out to be not significant when added to the final model with bare stem frequency (word-form frequency: t=0.047, p=0.96). Bare stem frequency remains significant in this model (t=-2.37, p=0.033).

In the model with relative frequency and word-form frequency, word-form frequency is only marginally significant at the point of its elimination (t=-1.83, p=0.09). At this point, relative

frequency is still significant (t=2.51, p=0.023) and only significant predictors remain in the model (POSITION, SPEECHRATE, NUMPHON, VOICERATIO). The direction of the two frequency effects is as predicted by H3a and H3c. Higher word-form frequency goes together with shorter stem duration (*coefficient*=-0.066, *std. error*=0.036, while higher relative frequency goes together with longer stem durations (*coefficient*=-0.076, *std. error*=0.030).

To address the collinearity issue further we employed random forest analysis (cf. Tomaschek, Hendrix & Baayen 2018) using conditional inference trees (`cforest()`, R package `party`, Hothorn et al. 2020). The random forest analysis showed that bare stem frequency is a more important predictor than relative frequency or word-form frequency. This supports our conclusions from the analyses presented in the previous paragraphs. Furthermore, bare stem frequency is among the most important predictors, in the same range as the baseline word duration. The random forest analysis is documented in detail in the supplementary material of this article.

5.2.2 Controlling for word-form frequency

To further disentangle the effects of bare stem frequency and word-form frequency we eliminated the potentially harmful correlation between bare stem frequency and word-form frequency by sampling words from a frequency band where there is no correlation between these two frequencies. This also eliminates the danger that the effect of bare stem frequency may be considered as being a hidden word-form frequency effect, due to the two being correlated.

We achieved this with the following procedure. Based on the inspection of the distribution of the two frequencies we first chose a narrow word-form frequency band in the middle of the distribution that had many observations. Figure 2 plots the two variables against each other. The size of the dots reflects the number of observations per word-form, as shown in the legend.

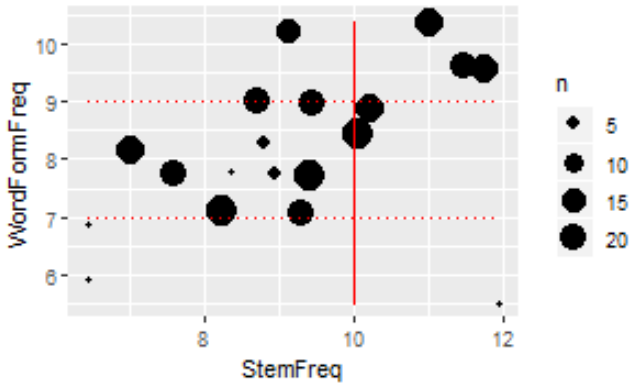

Figure 2: Plural word-form frequency by bare stem frequency (data set 2)

To narrow the range of word-form frequency we selected all observations with log word-form frequencies between 7 and 9, as indicated by the two dotted horizontal lines in Figure 2. To reduce the correlation of the two variables we further restricted the data set by selecting only those observations that had a log bare stem frequency of less than 10 (as indicated by the vertical red line in Figure 2). The resulting data set ('data set 3') covers the two central quartiles of word-form frequency and a few more data points below it (1st quartile: 7.11, median: 7.76, 3rd quartile: 8.62). This reduced data set still contains 159 observations (as against 295 in data set 2). The correlation of the three frequency variables in this data set are given in Table 5.

Table 5: Correlation matrix for lexical frequency measures in data set 3 (rho-values, p-values are given in parentheses, Spearman test)

	RELATIVEFREQ	STEMFREQ
WORDFORMFREQ	0.75 (0.000)	0.02 (0.79)
RELATIVEFREQ		-0.60 (0.000)

Bare stem frequency and word-form frequency can now be safely used in the same model. The final model is documented in Table 6.

Table 6: Fixed-effects coefficients and p-values in the final model testing the effects of STEMFREQ and WORDFORMFREQ on the duration of plural stems in data set 3. (Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05)

Random effects:					
Group	Name	Variance	Std. Dev.		
WORD	(Intercept)	0.0028	0.053		
Residual		0.064	0.25		
Fixed effects					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-1.92	0.52	5.31	-3.68	0.01
STEMFREQ	-0.10	0.03	4.31	-3.07	0.03
SPEECHRATE	-0.28	0.13	153.95	-2.16	0.03
NUMPHON	0.53	0.10	5.49	5.20	0.00
NEIGHBORDENSITY	0.01	0.01	4.61	2.39	0.07

The model shows a significant effect of bare stem frequency on the duration of the plural stem and no significant effect of word-form frequency (t=1.50, p=0.22 at the point of elimination). The latter was expected due to the narrow range of this variable. The analysis of this subset of data in which word-form frequency and bare stem frequency do not correlate thus lends strong support to H3b.

We also tested the effect of relative frequency in this data set (without also including word-form frequency, due to the high correlation of these variables). Relative frequency also had a significant effect on the duration of plural stems: Higher relative frequency goes together with longer duration of the stem, in accordance with H3c. Table 7 documents the final model.

Table 7: Fixed-effects coefficients and p-values in the final model testing the effect of RELATIVEFREQ on plural stem duration in data set 3. (Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05)

Random effects:					
Groups	Name	Variance	Std. Dev.		
WORD	(Intercept)	0.001542	0.03927		
Residual		0.063582	0.25216		
Fixed Effects					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-2.19	0.52	5.17	-4.22	0.01
RELATIVEFREQ	0.09	0.02	4.18	3.61	0.02
EXPWORDDUR	0.42	0.21	10.46	2.01	0.07
SPEECHRATE	-0.30	0.13	152.35	-2.32	0.02
NUMPHON	0.51	0.09	4.21	5.44	0.00
NEIGHBOURDENSITY	0.02	0.01	6.93	3.57	0.01

6. Discussion

In the present study we tested three predictions that follow from work on durational effects of paradigm uniformity such as Seyfarth et al. (2017). In contrast to Seyfarth et al. (2017), we used natural conversational speech instead of experimentally elicited speech, and we used data from New Zealand English instead of American English.

We repeat the three hypotheses investigated for the reader's convenience:

(1') H1: **Categorical paradigm uniformity**
Stems of plural words ending in the suffix [z] have longer durations than the pseudo-stems of mono-morphemic words ending in [z].

(2') H2: **Gradient paradigm uniformity due to activation strength** (Seyfarth et al. 2017)
a.) The higher the absolute frequency of the bare stem, the longer the duration of the plural stem.
b.) The higher the relative frequency, the shorter the stem of the inflected word.

(3') H3: **Gradient paradigm uniformity due to phonetic reduction**
a.) The higher the absolute frequency of the plural word-form, the shorter its duration, and the shorter the duration also of its stem (general reduction effect).

b.) The higher the absolute frequency of the bare stem, the shorter the duration of the plural stem.

c.) The higher the relative frequency (and keeping plural word-form frequency constant), the longer the duration of the plural stem.

We did not find the effect predicted by H1. Plural stems are longer than mono-morphemic pseudo-stems, but this difference is not significant if other variables are also taken into account. Like any null effect, the lack of a categorical paradigm uniformity effect needs to be interpreted with caution, as it may have different causes.

First, there may be a lack of statistical power. The fact that a number of other expected effects emerged as significant indicates that if there is also an effect of categorical paradigm uniformity, this effect is comparatively small.

Second, the absence versus presence of the effect might also be due to the kinds of data being used. Seyfarth et al. (2017)'s data contained a larger amount of tokens than our data set, but fewer word types, with only 16 plural forms and 16 homophones. The effect they found averaged over 16 plurals and 10 third person singular forms, and over final /s/ and /z/. The present study has used a smaller dataset (about half the size of Seyfarth et al. (2017)'s), which at the same time had many more different lexemes (40 lexemes with plural forms and 34 mono-morphemic lexemes, as against 16 pertinent homophone pairs in Seyfarth et al. (2017)'s experiment). The selection of the items from the corpus was highly restrictive (only monosyllabic words ending in a vowel-plus-/z/ sequence were used), but in comparison to a well-controlled experiment, the present data are still smaller, less balanced and more variable.

However, effects that have been observed under laboratory conditions need to be tested also under the conditions of everyday language use. If these tests fail outside the lab, the discrepancies are in need of explanations.

One explanation may be that previous studies compared frequency-matched members of homophone pairs in order to control for phonological make-up, while the present study compares non-matched non-homophones. Frequency-matched members of homophone pairs are very special, and they are a rather exceptional sample drawn from a much larger population of forms that are subject to very many different influences in production. A small effect in a sample with highly restricted properties may easily disappear in a less restricted sample.

But the null effect found in the present study may also indicate that the effect is spurious. Even in severely controlled samples of homophone pairs the categorical paradigm uniformity effect has not been consistently replicated. Seyfarth et al. (2017) themselves only find the effect for words ending in /s/ or /z/, but not for words ending in /d/ and /t/ (involving past tense as the

morphological category), even under experimental conditions. Other studies, such as Frazier (2006) or Seyfarth, Vander Klok & Garellek (2019) produced inconclusive or null results. Future studies of more phenomena in more languages are clearly called for to clarify the issues involved with the presence or absence of a possible categorical paradigm uniformity effect.

We also did not find a gradient uniformity effect based on the idea that greater activation strength of the bare stem leads to a stronger influence on the duration of the morphologically related form. In fact, we found the opposite direction of the effects predicted by that approach. The absence of the alleged effect is, however, expected, if we take well-known frequency-related phonetic reduction effects into account.

A higher frequency of the bare stem would go together with shorter duration. And if the influence on the duration of related forms is exerted via the articulatory plans, the opposite effect can be expected: The shorter duration of the high-frequency bare stem will lead to a shorter duration of the plural stem. The effects that follow from including effects of phonetic reduction are formulated in H3.

Hypothesis 3a is concerned with the general word-form frequency effect on duration. In the larger data set (data set 1) we found the predicted effect of word-form frequency as a main effect, with no interaction with MORPHEMETYPE. This means that the effect holds across the board, affecting plural words and mono-morphemic words in the same way, supporting H3a. The subset of the data that only contained the plural forms (data set 2), did not show a significant word-form frequency effect. These two findings together may well be an indication of a lack of statistical power for this rather small data set.

Hypotheses 3b and 3c are concerned with the gradient paradigm uniformity effect in relation to the frequencies of the forms involved. The predictions based on the hypotheses are fully supported by our results. Stems in plural words that are based on low frequency bare stems show longer durations, while plurals based on high frequency bare stems show shorter durations. This is fully in line with the idea that higher frequency of occurrence of a form leads to shorter durations. Following up on the idea that the stored articulatory gestures of a bare stem may influence the gestures used to produce morphologically related forms, these related forms show traces of the stem pronunciation, in this case its duration.

Recall that Seyfarth et al. (2017) found that neither the absolute frequency of the stem nor relative frequency had any effect on the duration of the inflected stems. How can the discrepancy between their results for gradient paradigm uniformity and ours be explained? Seyfarth et al. do not discuss their null result concerning gradient paradigm uniformity, but concede from the beginning that their "analyses should be interpreted with caution, in particular

because the stimuli were not selected to include a broad range of either frequency measure" (2017:9). Apparently, our corpus data had distributions that were such that they enabled effects to surface.

The effects of word-form frequency found in the present study are fully in line with those for lemma frequency (Jurafsky et al. 2001; Bell et al. 2009; Gahl 2008; Lohmann 2018) and those for word-form frequency (Caselli, Caselli & Cohen-Goldberg 2016; Lõo et al. 2018).

Particularly pertinent are the findings in Caselli, Caselli & Cohen-Goldberg (2016). These authors demonstrated that the bare stem frequency of words inflected with *-ed* and *-ing* negatively correlates with the duration of these words in speech. Although Caselli, Caselli & Cohen-Goldberg (2016) did not measure the stem duration of the inflected words (but the duration of the whole word instead), it can be safely assumed that not only the whole word, but also the stem showed this negative correlation between bare stem frequency and duration of the plural stem (see Plag, Homann & Kunter (2017) and Plag et al. (2020) for evidence and discussion of the general relation between word duration and stem duration).

The question of phonetic consequences of paradigm uniformity may be seen as part of a much larger set of questions concerned with the mutual influence of lexically related forms in speech production. Restricting ourselves to that part of the recent literature that focuses on acoustic properties, there is one study in particular that has looked at the effects of frequency on how similar words may influence each other's pronunciation (Goldrick et al. 2011). That study, like ours, tested conflicting predictions across pairs of forms of varying frequency. These authors investigated how in speech errors (like in the outcome *path* for intended target *bath*) the frequency of the target and of the erroneous outcome influence the phonetic properties of the outcome. They find that low frequency targets produce larger phonetic traces in the outcomes, and that low frequency outcomes are less influenced by phonetic traces of the target. These phonetic traces include vowel duration as a secondary cue to voicing. These results, like ours, are in line with other studies that have shown that low frequency words exhibit enhanced phonetic processing, resulting in, among other things, longer durations.

Other researchers have proposed that morphologically conditioned phonetic effects may arise from competition between language-specific general phonological patterns and word-specific structures, leading to intermediate forms or greater variability (e.g. Gafos (2006), Van Oostendorp (2008), Winter & Roettger (2011)). Consider the velarization of /l/ in English. The /l/ in suffixed forms like *knee.l#ing* is considerably darker than the /l/ in the same syllabic position of mono-morphemic words (Sproat & Fujimura 1993; Lee-Kim, Davidson & Hwang 2013). This may be explained as an effect of paradigmatic uniformity, such that the velarized

coda-/l/ of the stem influences the pronunciation of the onset-/l/ in the morphologically related suffixed form. Alternatively, the phonetic traces of the stem may arise through the competition between the general preference for clear [l] in onset position in English on the one hand, and the word-specific expectation for [ɬ] (cf. *kneel, kneels, knelt*, all featuring [ɬ]). However, the variability in the duration of stems does not involve a tension between general versus word-specific phonological patterns. This is therefore not a valid explanation for our findings.

The results of the present study, as well as the results of many other studies showing morpho-phonetic effects, are a challenge for modular phonological theories (such as Lexical Phonology (Kiparsky 2015), and for modular, strictly feed-forward models of speech production (such as the Levelt model in Levelt, Roelofs & Mayer (1999)). In these models the morphological information is no longer available at post-lexical stages. Paradigm uniformity effects thus seem to provide prima facie evidence for interactive models of speech production, in which spreading activation of morphologically related words plays a prominent role (e.g Ernestus & Baayen 2006; Ernestus & Baayen 2007; Baayen et al. 2007; Roettger et al. 2014; Dell 1986; Goldrick 2006; Goldrick 2014).

Funding Information

We are very grateful to the Deutsche Forschungsgemeinschaft for funding this research in the context of the DFG-Research Unit FOR2373 'Spoken Morphology' (Grants: PL151/8-2 ‘Morpho-phonetic Variation in English’ to Ingo Plag and PL151/7-2 ‘FOR 2737 Spoken Morphology: Central Project’ to Ingo Plag).

Acknowledgements

We are thankful to the two anonymous reviewers and the editors for their very helpful comments on a previous version of this paper. The paper also profited from the feedback of the members of the audiences at the following conferences: 12th Mediterranean Morphology Meeting (Ljubljana), 15. Phonetik und Phonologie Tagung (Düsseldorf) and International Morphological Processing Conference 2019 (Tübingen). We would also like to thank our colleagues in the DFG-Research Unit FOR2373 'Spoken Morphology' for their feedback on this project.

References

Baayen, R. Harald. 2008. *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. Harald, W. Levelt, Robert Schreuder & Mirjam Ernestus. 2007. Paradigmatic structure in speech production. In *Proceedings from the annual meeting of the Chicago linguistic society*, vol. 43, 1–29. Chicago Linguistic Society.

Baayen, R. Harald & Elnaz Shafaei-Bajestan. 2019. *languageR: Analyzing Linguistic Data: A Practical Introduction to Statistics*. https://CRAN.R-project.org/package=languageR (20 April, 2020).

Barton, Kamil. 2009. MuMIn: multi-model inference. *http://r-forge. r-project. org/projects/mumin/*.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2015. *The Oxford Reference Guide to English Morphology*. Oxford University Press.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 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, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*. Acoustical Society of America 113(2). 1001–1024.

Bell, Melanie J., Sonia Ben Hedia & Ingo Plag. 2019. How morphological structure affects phonetic realization in English compound nouns.

Ben Hedia, Sonia. 2019. *Gemination and degemination in English affixation: Investigating the interplay between morphology, phonology and phonetics*. Studies in Laboratory Phonology. https://doi.org/10.5281/zenodo.3232849 (30 September, 2019).

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

Bethin, Christina Y. 2012. On paradigm uniformity and contrast in Russian vowel reduction. *Natural Language & Linguistic Theory* 30(2). 425–463.
Blazej, Laura J. & Ariel M. Cohen-Goldberg. 2015. Can We Hear Morphological Complexity Before Words Are Complex? *Journal of Experimental Psychology: Human Perception and Performance* 41(1). 50–68. https://doi.org/10.1037/a0038509.
Boersma, Paul & David Weenink. 2015. *Praat: doing Phonetics by Computer*. (Version 6.0.08). http://www.fon.hum.uva.nl/praat/.
Bonami, Olivier, Gilles Boyé, Matthew Baerman, Oliver Bond & Andrew Hippisley. 2019. Paradigm uniformity and the French gender system. *Perspectives on morphology. Edinburgh: Edinburgh University Press, to appear*.
Braver, Aaron. 2014. Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping. *Lingua* 152. 24–44.
Burzio, Luigi. 1998. Multiple correspondence. *Lingua*. Elsevier 104(1–2). 79–109.
Byrd, D., J. Krivokapic & S. Lee. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *Journal Of The Acoustical Society Of America* 120(3). 1589–1599. https://doi.org/10.1121/1.2217135.
Byrd, Dani. 1994. Relations of sex and dialect to reduction. *Speech Communication* 15(1–2). 39–54. https://doi.org/10.1016/0167-6393(94)90039-6.
Caselli, Naomi K., Michael K. Caselli & Ariel M. Cohen-Goldberg. 2016. Inflected words in production: Evidence for a morphologically rich lexicon. *The Quarterly Journal of Experimental Psychology* 69(3). 432–454.
Cohen, Clara. 2014. Probabilistic reduction and probabilistic enhancement. *Morphology* 24(4). 291–323. https://doi.org/10.1007/s11525-014-9243-y.
Dell, Gary S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological review*. American Psychological Association 93(3). 283.
Eddington, David. 2006. Paradigm uniformity and analogy: The capitalistic versus militaristic debate. *International Journal of English Studies* 6(2). 1–18.
Engemann, U. Marie & Ingo Plag. 2020. Paradigm uniformity effects in spontaneous speech. *submitted to The Mental Lexicon*.
Engemann, U. Marie, Ingo Plag & Julia Zimmermann. 2019. Paradigmatic effects in speech production: Do bare stems influence the pronunciation of suffixed forms? In *MoProc 2019 - International Morphological Processing Conference*. Tübingen, Germany.
Ernestus, Mirjam & Harald Baayen. 2007. Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch. *Language and Cognitive Processes*. Routledge 22(1). 1–24. https://doi.org/10.1080/01690960500268303.
Ernestus, Mirjam & R. Harald Baayen. 2006. The functionality of incomplete neutralization in Dutch: The case of past-tense formation. (Ed.) L. Goldstein, D.H. Whalen & C.T. Best. *LabPhon* 8. 27–49.
Fougeron, C. & P. A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101(6). 3728–3740.
Fox, John & Sanford Weisberg. 2011. Multivariate linear models in R. *An R Companion to Applied Regression. Los Angeles: Thousand Oaks*.
Frazier, Melissa. 2006. Output-output faithfulness to moraic structure: Evidence from American English. In *PROCEEDINGS-NELS*, vol. 36, 1.
Fromont, Robert & Jennifer Hay. 2012. LaBB-CAT: an Annotation Store. In *Proceedings of Australasian Language Technology Association Workshop*, 113--117. Australasian Language Technology Associatio. http://labbcat.sourceforge.net/ (6 May, 2019).
Gafos, Adamantios I. 2006. Dynamics in grammar: Comment on Ladd and Ernestus & Baayen* Adamantios I. Gafos. *Laboratory phonology* 8(4). 51.

Gafos, Adamantios I. & Angela Ralli. 2002. Morphosyntactic features and paradigmatic uniformity in two dialectal varieties of the island of Lesvos. *Journal of Greek linguistics* 2(1). 41–73.

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

Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66(4). 806. https://doi.org/10.1016/j.jml.2011.11.006.

Goldrick, Matthew. 2006. Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*. Routledge 21(7–8). 817–855. https://doi.org/10.1080/01690960600824112.

Goldrick, Matthew. 2014. Phonological processing: The retrieval and encoding of word form information in speech production. In *The Oxford handbook of language production*, 228–244. Oxford, UK: Oxford University Press.

Goldrick, Matthew & Sheila E. Blumstein. 2006. Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*. Taylor & Francis 21(6). 649–683.

Goldrick, Matthew, H. Ross Baker, Amanda Murphy & Melissa Baese-Berk. 2011. Interaction and representational integration: Evidence from speech errors. *Cognition* 121(1). 58–72. https://doi.org/10.1016/j.cognition.2011.05.006.

Green, Christopher R. 2009. Paradigm uniformity in Luwanga derived nouns. In *6th World Congress on African Linguistics, Cologne, Germany. August*, 17–21.

Hay, Jennifer. 2003. *Causes and Consequences of Word Structure* (Outstanding Dissertations in Linguistics). Psychology Press.

Hay, Jennifer. 2007. The phonetics of 'un.' *Lexical creativity, texts and contexts* 39–57.

Hothorn, Torsten, Kurt Hornik, Carolin Strobl & Achim Zeileis. 2020. *party: A Laboratory for Recursive Partytioning*. https://CRAN.R-project.org/package=party (20 April, 2020).

Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In *Frequency and the emergence of linguistic structure* (Typological Studies in Language, Vol. 45), 229–254. Amsterdam, Netherlands: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.13jur.

Kemps, Rachel J. J. K., Mirjam Ernestus, Robert Schreuder & R. Harald Baayen. 2005. Prosodic cues for morphological complexity: the case of Dutch plural nouns. *Memory & Cognition* 33(3). 430.

Kenstowicz, Michael & Hyang-Sook Sohn. 2008. Paradigmatic uniformity and contrast: Korean liquid verb stems. *Phonological Studies* 11. 99–110.

Kiparsky, Paul. 2015. Stratal OT: A Synopsis and FAQs. In *Capturing phonological shades within and across languages*, 2–44. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59(5). 1208–1221. https://doi.org/10.1121/1.380986.

Labov, William. 1972. *Sociolinguistic Patterns*. University of Pennsylvania Press.

Laks, Lior, Evan-Gary Cohen & Stav Azulay-Amar. 2016. Paradigm uniformity and the locus of derivation: The case of vowel epenthesis in Hebrew verbs. *Lingua* 170. 1–22. https://doi.org/10.1016/j.lingua.2015.10.004.

Lee-Kim, Sang-Im, Lisa Davidson & Sangjin Hwang. 2013. Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology* 4(2). 475–511. https://doi.org/10.1515/lp-2013-0015.

Levelt, Willem J M, Ardi Roelofs & Antje S. Mayer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22(1). https://www.researchgate.net/publication/27269195_A_theory_of_lexical_access_in_speech_production.

Lohmann, Arne. 2017. Phonological properties of word classes and directionality in conversion. *Word Structure*. Edinburgh University Press The Tun-Holyrood Road, 12 (2f) Jackson's Entry … 10(2). 204–234.

Lohmann, Arne. 2018. Cut (n) and cut (v) are not homophones: Lemma frequency affects the duration of noun–verb conversion pairs. *Journal of Linguistics* 54(4). 753–777. https://doi.org/10.1017/S0022226717000378.

Lõo, Kaidi, Juhani Järvikivi, Fabian Tomaschek, Benjamin V. Tucker & R. Harald Baayen. 2018. Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology* 28(1). 71–97. https://doi.org/10.1007/s11525-017-9318-7.

Mackenzie, Sara, Erin Olson, Meghan Clayards & Michael Wagner. 2018. North American/l/both darkens and lightens depending on morphological constituency and segmental context. *Laboratory Phonology*. Ubiquity Press 9(1).

Marian, Viorica. 2012. *CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities*. United States, North America: Public Library of Science (PLoS).

McMillan, Corey T., Martin Corley & Robin J. Lickley. 2009. Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*. Routledge 24(1). 44–66. https://doi.org/10.1080/01690960801998236.

Park, Sunwoo. 2006. *Paradigm uniformity effects in Korean phonology*. PhD dissertation, Korea University, Seoul, Korea.

Peterson, RR & P Savoy. 1998. Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology*.

Plag, Ingo & Sonia Ben Hedia. 2018. The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration.

Plag, Ingo, U. Marie Engemann & Gero Kunter. 2018a. The effect of morphological boundaries on stem vowel duration in English. In *40. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft*. Stuttgart: Deutsche Gesellschaft für Sprachwissenschaft.

Plag, Ingo, U. Marie Engemann & Gero Kunter. 2018b. The effect of morphological boundaries on stem vowel duration in English. In *LabPhon 16 - Variation, development and impairment: Between phonetics and phonology*. Lisbon: Association for Laboratory Phonology.

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

Plag, Ingo, Arne Lohmann, Sonia Ben Hedia & Julia Zimmermann. 2020a. An <s> is an <s'>, or is it? Plural and genitive-plural are not homophonous. In *To appear in Livia Körtvélyessy & Pavel Stekauer (eds.) Complex Words*. Cambridge, UK: Cambridge University Press.

Plag, Ingo, Arne Lohmann, Sonia Ben Hedia & Julia Zimmermann. 2020b. What is the difference between _boys_ and_boys'_? The phonetics of plural vs. genitive-plural in English and its implications for morphological theory. In *19th International Morphology Meeting*. Vienna University of Economics and Business, Vienna, Austria.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005a. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*. Karger Publishers 62(2–4). 146–159.

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

Pluymaekers, Mark, Mirjam Ernestus, R. Harald Baayen & Geert Booij. 2010. Morphological effects on fine phonetic detail: The case of Dutch-igheid. (Ed.) C Fougeron, B Kühnert, M D'Imperio & N Vallée. *Laboratory phonology* 10. 511–531.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. (Version 3.2.1). Vienna, Austria. https://www.R-project.org.

Raffelsiefen, Renate. 2004. Paradigm Uniformity Effects Versus Boundary Effects. In *Paradigms in Phonological Theory*. Oxford, UK: Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199267712.001.0001/acprof-9780199267712-chapter-9 (10 April, 2019).

Ramig Lorraine A. & Ringel Robert L. 1983. Effects of Physiological Aging on Selected Acoustic Characteristics of Voice. *Journal of Speech, Language, and Hearing Research*. American Speech-Language-Hearing Association 26(1). 22–30. https://doi.org/10.1044/jshr.2601.22.

Rapp, B & M Goldrick. 2000. Discreteness and interactivity in spoken word production. *Psychological review*.

Rebrus, Péter & Miklós Törkenczy. 2005. *Uniformity and contrast in the Hungarian verbal paradigm*. na.

Riehl, Anastasia K. 2003. American English flapping: Perceptual and acoustic evidence against paradigm uniformity with phonetic features. *Working Papers of the Cornell Phonetics Laboratory* 15(271–337).

Roettger, T. B. 2014. Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43. 11.

Roettger, Timo B., Bodo Winter, S. Grawunder, J. Kirby & M. Grice. 2014. Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43. 11–25. https://doi.org/10.1016/j.wocn.2014.01.002.

Schmitz, Dominic, Ingo Plag & Dinah Baer-Henney. 2020. How real are acoustic differences between different types of final /s/ in English? Evidence from pseudowords. In *19th International Morphology Meeting*. Vienna University of Economics and Business, Vienna, Austria.

Seyfarth, Scott, Marc Garellek, Gwendolyn Gillingham, Farrell Ackerman & Robert Malouf. 2017. Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience* 33(1). 32–49.

Seyfarth, Scott, Jozina Vander Klok & Marc Garellek. 2019. Evidence against interactive effects on articulation in Javanese verb paradigms. *Psychonomic bulletin & review* 1–7.

Skoog Waller, Sara, Mårten Eriksson & Patrik Sörqvist. 2015. Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in Psychology*. Frontiers 6. https://doi.org/10.3389/fpsyg.2015.00978. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00978/full (22 April, 2020).

Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English/l/and its implications for phonetic implementation. *Journal of phonetics* 21(3). 291–311.

Steriade, Donca. 2000. Paradigm Uniformity and the Phonetics-Phonology Boundary. (Ed.) Edited Michael Broe & Janet Pierrehumbert. *Papers in Laboratory Phonology* 5.

Tabain, Marija. 2003. Effects of prosodic boundary on /aC/ sequences: articulatory results. *The Journal of the Acoustical Society of America* 113(5). 2834–2849.

Tomaschek, Fabian, Peter Hendrix & R. Harald Baayen. 2018. Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics* 71. 249–267. https://doi.org/10.1016/j.wocn.2018.09.004.

Tomaschek, Fabian, Ingo Plag, Mirjam Ernestus & R. Harald Baayen. 2019. Modeling the duration of word-final S in English with Naive Discriminative Learning. *submitted to Journal of Linguistics*.

Torreira, Francisco & Mirjam Ernestus. 2009. Probabilistic effects on French [t] duration. In *10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, 448–451. Causal Productions Pty Ltd.

Tucker, Benjamin V. & Mirjam Ernestus. 2016. Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The mental lexicon*. John Benjamins 11(3). 375–400.

Van Oostendorp, Marc. 2008. Incomplete devoicing in formal phonology. *Lingua*. Elsevier 118(9). 1362–1374.

Walsh, Liam, Jen Hay, Derek Bent, Jeanette King, Paul Millar, Viktoria Papp & Kevin Watson. 2013. The UC QuakeBox Project: Creation of a community-focused research archive. https://ir.canterbury.ac.nz/handle/10092/15635 (20 November, 2018).

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America* 91(3). 1707–1717. https://doi.org/10.1121/1.402450.

Winter, Bodo & Timo B. Roettger. 2011. The nature of incomplete neutralization in German: Implications for laboratory phonology. *Grazer Linguistische Studien* 76. 55–74.

Zee, Tim. 2019. Morphological effects on the acoustics of Dutch /s/. In *15. Phonetik und Phonologie Tagung*. Düsseldorf, Germany.

Zimmermann, Julia. 2016. Morphological Status and Acoustic Realization: Findings from NZE. In C Carignan & M.D. Tyler (eds.), *Proceedings of the 16th Australasian International Conference on Speech Science and Technology*, 6–9. Sydney: University of Western Sydney.

Zuraw, Kie & Sharon Peperkamp. 2015. Aspiration and the gradient structure of English prefixed words. In *ICPhS*.

Appendix

Types and number of tokens

Table 8: Types and number of tokens, mono-morphemic words are shown on the left, and plural words on the right side of the table.

mono-morphemic	**number of observations**	**plural**	**number of observations**
bruise	3	bars	5
cause	5	boys	20
cheese	4	cars	21
chose	5	days	21
close	21	doors	24
daze	4	eyes	16
ease	3	floors	17
froze	5	guys	24
hose	4	jars	3
lose	13	keys	24
noise	20	knees	18
phase	3	laws	4
raise	3	news	23
rise	7	shoes	23
rose	4	stars	5
size	23	stores	3
vase	3	toes	3
wise	6	trees	20
		twos	3
		ways	18

Random forest analysis

Table 9: Variables and their importance values in the random forest analyses

Variable	Importance	Variable	Importance
WORD	0.0268	WORD	0.0260
EXPWORDDUR	0.0092	EXPWORDDUR	0.0090
STEMFREQ	0.0082	STEMFREQ	0.0079
NEIGHBORDENSITY	0.0047	NEIGHBORDENSITY	0.0046
VOICERATIO	0.0036	NUMPHON	0.0034
NUMPHON	0.0028	VOICERATIO	0.0034
RELATIVEFREQ	0.0028	RELATIVEFREQ	0.0024
WORDFORMFREQ	0.0019	WORDFORMFREQ	0.0019
SPEECHRATE	0.0014	SPEECHRATE	0.0011
POSITION	0.0011	POSITION	0.0010
NEIGHBORFREQ	0.0009	NEIGHBORFREQ	0.0008
LBIGRAMPROB	0.0006	LBIGRAMPROB	0.0003
AGEGROUP	0.0002	AGEGROUP	0.0002
RBIGRAMPROB	-0.0002	RBIGRAMPROB	-0.0003