AN <S> IS AN <S'>, OR IS IT?

PLURAL AND GENITIVE-PLURAL ARE NOT HOMOPHONOUS

Ingo Plag, Sonia Ben Hedia, Arne Lohmann & Julia Zimmermann

1. Introduction

Recent research on the acoustic properties of morphologically complex words has shown unexpected effects of morphology on phonetic realization. For instance, it has been demonstrated that homophonous suffixes such as final S in English may differ systematically in their phonetic properties (e.g. Zimmermann 2016, Plag et al. 2017, Seyfarth et al. 2017, Tomaschek et al. 2018). And even a particular kind of final S, i.e. third person singular, has been shown to vary phonetically according to morphological properties, such paradigmatic probability (Cohen 2014).

Such findings are unexpected since standard feed-forward theories of morpho-phonology (e.g. Lexical Phonology) and of speech production (e.g. Levelt, Roelofs & Meyer 1999) exclude such effects because there is no mechanism that would allow morphology to influence articulation, or that would model such behavior. English final S is an interesting test case in this domain that has received some attention in recent years.

In the spontaneous American English speech as collected in the Buckeye Corpus (Pitt et al. 2007) non-morphemic S is longer in duration than suffix S, and suffix S is longer than the S resulting from cliticization of *has* or *is* (Plag et al. 2017, Tomaschek et al. 2019). Similar results have been obtained with data from New Zealand English (Zimmermann 2016). For some types of S, especially for genitive-plural, little is known about their phonetic properties, since these forms have either not been investigated at all, or, as in the corpus-based studies just mentioned, the sample of these forms was too small to draw any firm conclusions.

In this paper we focus on the potential durational contrast between plurals and genitive-plurals (as in *boys* vs. *boys'*), i.e. of two forms that are standardly assumed to show no systematic phonological or phonetic difference (cf., for example, Zwicky 1975, Bauer et al. 2013: 145). We report the results of an experiment in which sentences were read aloud which contained pertinent words in very similar contexts (see Lohman & Conwell 2017, who provided the original data set).

The 462 plural tokens and the 417 genitive-plural tokens gleaned from the experiment were phonetically annotated, and the duration of S as well as the duration of the whole word was analyzed using mixed effects regression models with pertinent co-variates (e.g. speech rate, voicing, lexical frequency etc.). The results show that plural S is significantly shorter than genitive-plural S, with a mean difference of 7 to 8 ms between plural S and genitive-plural S (as predicted by different regression models). The duration effect is, however, not restricted to the final S, but extends over the whole word, with (monosyllabic) plural nouns being 14 ms shorter than genitive-singular nouns. The paper concludes by discussing how the present result can be explained in the light of morphological theories and various approaches to speech production.

2. Final S in English

Traditionally, phonetics plays no role in morphology. While morphological structure may affect phonological structure, it is not thought to influence phonetic detail. For instance, the standard literature (e.g. Bauer et al. 2013: chapter 1, Palmer et al. 2002) holds that there are three different allomorphs of the regular plural S: /ɪz/ after sibiliants, /z/ after voiced sounds, /s/ after unvoiced sounds. According to Bauer et al. (2013: 15) "[t]his allomorphy is easily understood in phonological terms (assimilation and epenthesis to break up illegal geminates), and is not controversial". The regular genitive-plural has exactly the same allomorphs, with the complication that these allomorphs are the exponents of two morpho-syntactic features at the same time, plural and genitive. This phenomenon may therefore be analyzed as a case of cumulative exponence. Alternatively one could assume that only one of the two features is overtly expressed. This seems to be the view held by people who call the genitive-plural 'bare genitive'. According to this analysis of the genitive-plural, "[i]n speech it [the genitive] has no realisation at all, such genitives being identical with the non-genitive" (Palmer et al. 2002: 1595). In writing, the genitive feature is represented by an apostrophe following the plural <s> (as in *boys', dogs'*, Bauer et al. 2013:144f).

Interestingly, irregularly inflected plural nouns like *geese* and *mice*, although ending in /s/, express the genitive feature with the allomorph that regularly follows stem-final non-morphemic /s/, i.e. /ɪz/: *geese's* /gisɪz/, *mice's* /maɪsɪz/. The allomorphy of the genitive-plural therefore depends on the morphological status of the final sibilant, i.e. the presence of the plural S is necessary for the occurrence of the bare genitive in genitive-plurals. In what follows, when talking about 'genitive-plural' forms we restrict ourselves to regularly inflected plural nouns that also carry a genitive feature. In other words, nouns like *geese* or *mice* are excluded.

The view that in speech the plural and the genitive-plural are identical, or that "as spoken, /dɒgz/ is ambiguous between genitive singular *dog's*, non-genitive-plural *dogs*, and genitive-plural *dogs'* " (Palmer et al. 2002: 1595), may, however be wrong. Recent research in morpho-phonetics has revealed that morphological information may impact fine phonetic detail. In particular, phonologically homophonous morphological units may exhibit systematic acoustic or articulatory differences. For instance, Kemps et al. (2005) and Blazej & Cohen-Goldberg (2015) have shown that free and bound variants of a base differ in duration, and Tomaschek et al. (submitted) demonstrate that articulatory movements of verbal stems differ systematically between suffixed and unsuffixed verbs. With regard to final S in English, speech corpus studies of North American and New Zealand English have found differences in duration between different kinds of S (non-morphemic, suffixal and auxiliary clitic S, Zimmermann 2016, Plag et al. 2017, Tomaschek et al. 2019). Some of the observed durational differences are quite large (e.g. 47 milliseconds between the observed means of non-morphemic S and the *has* clitic, Plag et al. 2017: 208). These studies have also included tokens of the genitive-plural, but this morphological form is too infrequent in the available corpus data to allow for firm conclusions. Furthermore, the phonetic properties of genitive-plural nouns have not been investigated in experiments yet. Different predictions are possible for the behavior of the two categories plural and genitive-plural, depending, among other things, on the theoretical analysis of the morpho-phonology of final S.

There are basically two formal approaches to the genitive-plural. One is what we will call 'selection', i.e. the selection of an exponent suffix (if we think in terms of morphemes) or of an inflected word-form (if we are more inclined towards a word and paradigm approach). In the selection approach the morpho-syntactic feature bundle <genitive, plural> is realized by the same forms as the feature <plural>, i.e. by /z/, /s/ or /ɪz/, and the correct form is chosen by the same mechanisms as the plural form. In the word and paradigm variant of the selection approach we would not select the suffix, but the word form that ends in the correct allomorph and has the correct morpho-syntactic specification. (1) illustrates this approach, giving also the exponents for the feature <genitive> (which has an additional exponent, ø, which may occur with proper nouns, e.g. *Burns'*, or in set expressions, e.g. *for goodness' sake*) .

(1)	feature specification	exponents
	<plural>	/z/, /s/ or /ɪz/
	<genitive>	ø, /z/, /s/ or /ɪz/
	<genitive, plural>	/z/, /s/ or /ɪz/

Under this approach there is no reason to expect a difference in phonetic realization between the plural and the plural-genitive.

The second approach involves haplology. According to this approach, some phonological material is not expressed due to a mechanism that avoids the expression of identical adjacent material (e.g. Plag 1998). In the case of the genitive-plural one could assume, at some level of representation, the presence of two exponents, one for plural and one for genitive. One of the two does not surface, for example due to a constraint against having geminated consonants, in this case two adjacent sibilants (e.g. *SIB-SIB, Russel 1997: 122f.).[1] The effect of this would be the same as in the selection approach, the complete deletion of one of the two segments, with the result that there is no phonetic difference between plural and plural-genitive.

To summarize, from a traditional structural point of view we can formulate the following hypothesis about the duration of plural S and genitive-plural S, which is at the same time our null hypothesis. We call this hypothesis the 'Structural Hypothesis':

(2) H0 'Structural Hypothesis'

There is no difference in duration between the plural S and the genitive-plural S.

The literature on speech production offers a number of alternative hypotheses relating to spelling, prosody lexical frequency and processing. We will discuss each in turn, starting with spelling. Brewer (2008) has demonstrated that, for literate speakers of English, speech production is not independent of orthography. In particular, sounds or sound sequences that have longer orthographic representations tend to have longer acoustic durations. For instance, the /f/ in *raft, staffed, graphed* corresponds to one or two letters. Brewer shows that there is a positive correlation between the number of letters and acoustic duration, even though the

[1] There is a complication that arises from the genitive-plural of stems taking the /ɪz/ allomorph for the plural, e.g. *horses*. These nouns also take the bare plural. In a selection account this not a problem, as the correct exponent is selected based on the final segment of the base. In a haplology account, the actual mechanisms would depend on how the alternation between the three exponents is generally accounted for. To solve the problem one may assume an underlying /z/ from which all exponents are derived.

underlying segment is the same. In addition to the amount of orthographic material, frequency also plays a role. The most frequent orthographic representations of sounds show the shortest durations. The effects are consistent in experiments and in the speech corpus used by Brewer (Buckeye), so that the effects cannot be dismissed as artefacts of online reading under experimental conditions.

As mentioned in section 2, the genitive-plural has one orthographic sign (the apostrophe) more than the plural. In terms of frequency, word-final <s'> is a very rare orthographic representations of /s/ or /z/, as it only occurs in plural-genitives and in proper nouns ending in <s>. In contrast, <s> is the most common orthographic representation of /s/ and /z/. The longer duration of the genitive-plural S could therefore be interpreted as another instance of the kinds of orthographic effect on speech production that Brewer detected. Based on Brewer's findings we can set up the second hypothesis:

(3) H1 'Spelling Hypothesis'
Plural S is shorter than genitive-plural S

Prosody also comes into play. It is well known that segments preceding prosodic boundaries are lengthened, with the amount of lengthening reflecting the strength of the prosodic boundary (e.g. Wightman et al. 1992). A difference in prosodic boundary strength following the two Ss may thus be another relevant factor resulting in a durational difference between the two. Most theories explaining prosodic boundary placement and strength rely to a considerable degree on the syntactic constituent structure of the sentence (see Turk & Shattuck-Hufnagel 2014 for an overview). While syntactic structure and prosodic structure are not isomorphic, syntactic and prosodic boundaries nevertheless tend to co-occur. In the target sentences of the present dataset the genitive-plural S occurs always phrase-medially, being embedded in a noun phrase (e.g. [*the* ***patients'*** *nap*]$_{NP}$), while the plural S occurs always in phrase-final position of an NP that precedes a VP (e.g. [*the* ***patients***]$_{NP}$ [*nap...*] $_{VP}$). This difference in position within the embedding syntactic constituent would predict a stronger prosodic boundary after plural S and consequently greater domain-final lengthening of this S. We can thus formulate the third hypothesis:

(4) H2 'Prosody hypothesis'
Plural S is longer than genitive-plural S

The two Ss in question are characterized by a considerable difference in usage frequency, with the plural S outnumbering the genitive-plural S by far. It is well-known that words are phonetically reduced, i.e. pronounced shorter, with increasing frequency (see Jurafsky et al. 2001, Gahl 2008). There are different explanations as to how the reductive effect of frequency comes about (see e.g. the discussion in Gahl 2008). One account, put forth by Bell et al. (2009), is that differences in duration reflect the speed of retrieval from the mental lexicon. Low-frequency forms take longer to retrieve than high-frequency forms. The greater duration of the former may thus be a way to adjust for asynchronies between retrieval and articulation. While Bell et al. (2009) dealt with differences at the lexical level, the same mechanism may also be at work at the morphological level. There is in fact evidence that effects of frequency can be observed at the sub-lexical level. With regard to the plural suffix, Rose (2017) demonstrates that contextual predictability, measured in terms of how often the preceding word occurs before a plural noun, has an effect on the duration of plural S. Plurals that are more predictable according to this measure tend to have more reduced realizations of S. Given that plurals in general can be assumed to have higher frequencies and probabilities than genitive-plurals, it is expectable that plural S is shorter than genitive-plural S:

(5) H3 ‘Morpheme-based Frequency Hypothesis’
Plural S is shorter than genitive-plural S

In terms of processing, one could assume that the retrieval of a given genitive-plural form takes longer than the retrieval of its corresponding plural form not only for reasons of frequency. The genitive-plural arguably involves the activation of two morpho-syntactic features. The slower retrieval process of the genitive-plural S might result in a longer duration in comparison to the plural S. This leads to another hypothesis:

(6) H4 ‘Complexity Hypothesis’
Plural S is shorter than genitive-plural S

To complicate matters further, a potential durational difference between plural S and genitive-plural S may also come about through durational effects affecting the whole word. Recent work on the production of inflected words has demonstrated that the frequency of a given individual word-forms in its paradigm influences the duration of its phonetic realization. For instance, Caselli et al. (2015) find that word-form frequency predicts the duration of English words suffixed with *-ing, -ed*, and *-s*, which reveals that both root and word representations play a role in the production of inflected English words. A similar result was obtained for Estonian noun inflection by Lõo and colleagues (Lõo et al. 2017). In both studies higher word-form frequency goes together with shorter word duration. Under the assumption that final S participates in this word-form frequency effect, we can derive two related hypotheses:

(7) H5 'Word-based Frequency Hypothesis'

a. The duration of final S depends on the frequency of the word-form it is part of. The more frequent the word-form, the shorter the S.

b. The duration of the whole word (i.e. the duration not only of the final S) depends on the frequency of the word-form. The more frequent the word-form, the shorter the whole word.

The investigation of the Word-based Frequency Hypothesis entails that we do not restrict ourselves to the duration of final S, but also look at the duration of the whole word.

The different approaches and hypotheses do not all make different predictions. Hence, whatever the facts turn out to be, the patterning of the durations might not necessarily support a particular theory. We will see, however, that some approaches make wrong predictions, which suggests that the underlying theories are in need of revisions.

3. Methodology

3.1. Stimuli and procedure

The data for our study come from a study by Lohmann & Conwell (2019), in which the authors tested durational difference between nouns and verbs in North American English. The experimental items were constructed in a way that allowed us to also investigate the durational difference between plural S and genitive-plural S. In the experiment sentences were read aloud which contained pertinent words in very similar contexts.

There were two types of sentences, in one of which there were pairs of phonologically homophonous pairs of plural and genitive-plural forms. We use only the data from this sentence type, which is illustrated in (8). The 'noun sentence' elicited the noun (given in italics) whose duration was of interest to Lohmann & Conwell, the 'verb sentence' elicited the corresponding verb (also given in italics). Preceding their target noun or verb, we find the noun that is of interest for the present study, given in bold. Sentences were presented in two variants, one with an additional preposition phrase (given in parentheses in (8)), the other without this phrase.

(8) a. *Context:*

Mike and his team are very busy finishing up the report for the end of the quarter. They see that some of their co-workers in accounting do not seem to take their work seriously.

Noun sentence:

Their **colleagues'** *nap* in the cubicle (next to the busy hallway) upsets the hard-working employees.

Verb sentence:

Their **colleagues** *nap* in the cubicle (next to the busy hallway) and this upsets the hard-working employees.

b. *Context:*

Dr. Butler and Dr. Gonzales have moved their practice out of the city. Now, some of the older patients are very sleepy when they arrive at the cardiologists' new office.

Noun sentence:

The **patients'** *nap* in the waiting room (with the new furniture) irritates the doctors.

Verb sentence:

The **patients** *nap* in the waiting room (with the new furniture) and this irritates the doctors.

To control for potential influences of intervening variables, the two sentences in a pair differed only minimally from each other in terms of their syntactic structure and lexical material. In order to reduce effects of priming or repetition each participant read out only one of the two forms of a lexeme (i.e. either the plural form, or the genitive-plural form). The only exception to this is the lexeme *actor* which occurred in two different sentence pairs.

In the final data set, each participant provided between 3 and 7 plural forms and between 2 and 7 genitive-plurals. On average a participant read 5.9 plurals and 5.3 genitive-plurals. Four of the 82 participants had to be excluded due to frequent disfluencies in their recordings. A more detailed discussion of the stimuli and the recording procedure can be found in Lohmann

& Conwell (2019).

The final data set for the present study consists of all observations that contained a target stem that was tested in both a plural context and in a genitive-plural context. Target words were excluded in which the consonant following our target item was /s/ (e.g. *mothers'* in the context *the mothers' snack*), since these items did not allow for setting a clear boundary between the two words. Appendix A contains a list of the stimuli that are included for analysis in the present study (13 sentence pairs). Overall, 879 words entered our analysis. They represent 12 different plural−genitive-plural word pairs. Table 1 gives an overview of the target words.

Table 1: Target stems with their frequency in the data set (*N*=879)

target stem	**genitive-plural**	**plural**	**sum**
actor	66	76	142
boy	31	32	63
colleague	32	38	70
corporation	32	34	66
dog	35	29	64
grandparent	35	37	72
Henderson	34	36	70
hiker	35	32	67
kid	29	37	66
parent	34	36	70
patient	21	36	57
student	33	39	72
sum	417	462	879

3.2. Data preparation

First an automatic segmentation of the acoustic data was carried out with the help of the MAUS forced alignment software (Kisler, Reichel & Schiel, 2017). This automatic segmentation was then manually corrected by trained research assistants. The research assistants followed the same protocol as Plag et al. (2017) in their study of S, relying on cues in the waveform and the spectrogram. The manual annotation was done using Praat (Boersma & Weenink, 2016). A Praat script then extracted the acoustic measurements that we were interested in.

3.3 Statistical analysis: Predictors and modeling procedures

To test the hypotheses we conducted several linear mixed effects regression analyses, with the morphological category (MORPH, values: plural and genitive-plural), and word-form frequency as the predictors of interest. In order to test H0 to H4, we fitted models with the duration of S as the dependent variable and MORPH as the variable of interest. In order to test H5 the variable of interest was word-form frequency. For H5a the dependent variable was the duration of S, for H5b it was the duration of the whole word.

We extracted the word-form frequencies from the DVD version of the *Corpus of Contemporary American English* (COCA) (Davies, 2013), using the query tool *Coquery* (Kunter, 2016). We consider COCA an adequate source for the frequency counts because the data in this corpus come from the same variety of English as the speech data under investigation. Following standard procedures we log-transformed word-form frequency to reduce the potentially harmful effect of skewed distributions in linear regression models. The name of this variable is LOGWORDFORMFREQUENCY.

In addition to the predictors of interest, we also added some noise variables to control for known effects of certain phonetic parameters. These noise variables largely overlap with those used in other studies, e.g. Plag et al. (2017). Not all noise variables are used in all models. Which variables were included in which models will be explained as we go along.

- VOICING. Voiced fricatives are shorter than unvoiced ones (e.g. Klatt 1976). In order to categorize an S as either voiced or unvoiced we used the proportion of pitch pulses in the segment. The distribution of this measurement was bimodal, indicating a categorical distinction. Following Plag et al. (2017), an S was considered to be voiced if the PRAAT algorithm detected voicing in more than 75 percent of the overall duration of the segment (given as 'voiced frames' in Praat). We also tested an interaction between VOICING and MORPH, since Plag et al. (2017) had found such an interaction in their sample. This interaction was not significant in any of our models.
- LOGBASEDURATION. The articulation rate (or 'speech rate') has an obvious influence on the duration of individual segments. The duration of the base is one indicator of articulation rate. All other things being equal a shorter base duration indicates a faster articulation. For instance, Plag et al. (2017) found that the duration of S correlates with the duration of the base which means that final S participates in lengthening or shortening processes that affect the whole word. We log transformed the duration measurements and called this variable LOGBASEDURATION.

- SPEECHRATE. Another frequently used measurement for articulation rate is the number of segments divided by the duration of a relevant linguistic unit. Segment durations become shorter with increasing values of this measurement. For our purposes, we computed a measurement we called SPEECHRATE as the quotient of the number of segments and the duration of the base.
- NUMBEROFSYLLABLES. Words with more syllables may tend to have shorter durations of the individual segments (see Plag et al. 2017). We included the number of syllables of the citation form of the target word as a (factorial) covariate.
- NUMBEROFCONSONANTS. The more consonants there are in a consonant cluster, the shorter the individual segments (Klatt 1976). We therefore coded the number of consonants in the rhyme of the final syllable (which contained the S) of our target words.
- FOLLOWINGSEGMENT. According to, for example, Klatt (1976, see also Plag 2017), the segment following the S may influence the duration of S. We coded the kind of segment following the target word (with the values affricate, lateral, nasal, plosive).
- LOGLEMMAFREQUENCY. More frequent words are pronounced with shorter durations (see, for example, Jurafsky et al. 2001, Gahl 2008, for a summary of the literature). We used the log-transformed lemma frequencies from COCA (Davies, 2008).
- GENDER. Some studies have found gender-related variation in speech rates of individual speakers (see van Borsel 2008 for an overview). We included the gender of the speaker as a co-variate.

The regression models were fitted using the packages lme4 (Bates et al. 2014) and lmerTest (Kuznetsova et al. 2017) for R (R Development Core Team 2014). We started with maximal models that contained a maximal reasonable subset (see section 4) of the above predictor variables as fixed effects plus random intercepts for subject and item (i.e. the lemma). To test for speaker-specific or item-specific effects of MORPH and LOGWORDFORMFREQUENCY, we also included a random contrast for MORPH by subject and one for MORPH by ITEM, as well as a random slope for LOGWORDFORMFREQUENCY by SUBJECT and a random slope for LOGWORDFORMFREQUENCY by ITEM.[2]

Following standard stepwise elimination procedures (e.g. Baayen 2008) a variable (fixed or random) was kept in the model if its inclusion led to a decrease in the AIC and to a significant improvement (p<0.05) in model fit tested via a log-likelihood test. To ensure a

[2] Speakers may vary in their sensitivity to the durational variation arising from morphological structure. For instance, Zimmermann et al. (2017) present data from two speakers, one of which showed a consistent durational contrast between plural S and *has*-clitic S, while the other does not.

consistent elimination procedure we used the `step` function of the `MASS` package (Venables & Ripley 2002).

3.4 Collinearity

One issue we had to address is collinearity (e.g. Tomaschek et al. 2018). Of our noise variables LOGBASEDUR correlated highly (and expectedly) with NUMBEROFSYLLABLES (*rho*=0.78, *p*<0.001, Spearman). We addressed this problem by not including LOGBASEDUR as a predictor in our models, knowing that this measurement was also used in the computation of SPEECHRATE.

SPEECHRATE still correlated with NUMBEROFCONSONANTS (*rho*=-0.65 *p*<0.001, Spearman), and a bit less strongly with NUMBEROFSYLLABLES (*rho*=-0.46, *p*<0.001, Spearman). These correlations were also expected since SPEECHRATE was computed with the number of segments in the base, and the number of segments in the base obviously correlates with the number of syllables and the number of consonants in the rhyme of the last syllable. To address this collinearity issue, we used each of the three predictors individually in the initial models, with the result that only SPEECHRATE turned out to be a statistically significant predictor of S duration. In the models for word duration both SPEECHRATE and NUMBEROFSYLLABLES were significant predictors. We calculated variance inflation factors for models with both variables, with the factors being 1.4 (NUMBEROFSYLLABLES) and 1.02 (SPEECHRATE), which indicates a very low danger of collinearity. We therefore included both variables in the word duration models.[3]

Furthermore, there was a very strong correlation between LOGLEMMAFREQUENCY and LOGWORDFORMFREQUENCY (*rho*=0.68, *p*<0.001). In the models in which LOGWORDFORMFREQUENCY was the variable of interest we therefore did not include LOGLEMMAFREQUENCY.

The two morphological categories vary significantly by LOGWORDFORMFREQUENCY, with the plural word forms having a much higher frequency than the genitive-plural forms (means: 9.5 vs 5.9, *W=23373, p*<2.2e-16, Wilcoxon test). Figure 1 shows the distribution by lexeme. Each pair of dots connected by a line represents one lexeme, with its two forms.

---

[3] We included speech rate in the model of word duration, although it is computed by including base duration. Base duration indeed correlates highly with word duration (*rho*=0.97, *p*<2.2e-16, Spearman), but speech rate does not correlate significantly with word duration (*rho*=0.02, *p*=0.61, Spearman). In models fitted to word duration that do not contain speech rate, the significant fixed effects are the same as in models with speech rate, but have larger coefficients. Including speech rate is therefore a conservative strategy that works against the significance of the variables of interest.

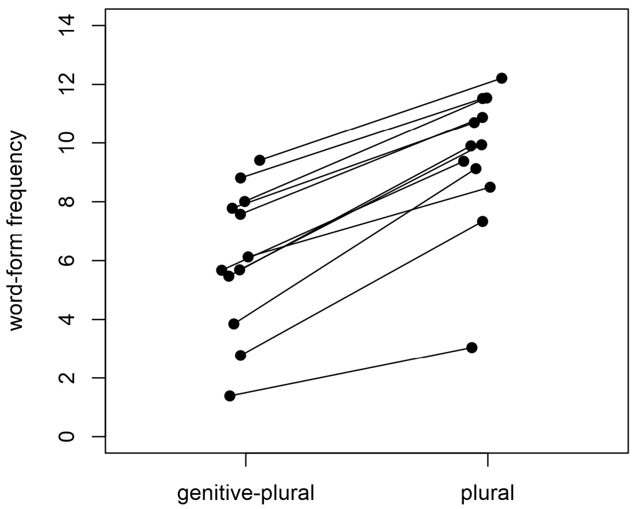


Figure 1: Word-form frequency by morphological category

Given that LOGWORDFORMFREQUENCY and MORPH co-vary significantly (see also *rho*= 0.66, *p*<2.2e-16, Spearman) precludes including them both in one model. We will therefore fit each model with only one of the two.

All models needed trimming of the residuals as the final stage the model fitting process. To ensure a satisfactory distribution of the residuals in the final models, we removed data points with residuals larger than 2.5 standard deviations. If this trimming was not enough, we removed data points with residuals larger than 2.0 standard deviations. This procedure led to a satisfactory distribution of the residuals in all models. The final regression models were based on very similar numbers of observations.

The final models are documented in Appendix B. The data set and the statistical modeling script are documented in full in the supplementary material for this article, which is available at https://osf.io/ubxgy/?view_only=29a47c7f66574f9385332fd68b8d6984.

4. The duration of plural S and genitive-plural S

4.1 Overview and data sets

Figure 2 shows the distributions of the observed durations of plural S and genitive-plural S in the untrimmed data set. On average, genitive-plural S is about 8 ms (or 10 percent) longer than plural S, with a mean of 74 ms duration for plural S and 82 ms for genitive-plural S. This difference is significant (Wilcoxon test, W=110710, p=0.00013). This may already be an interesting result, but given the many potentially intervening influences described in the previous section, these influences should be controlled for in a multivariate analysis, such as the mixed effects regression analysis described in the previous section.

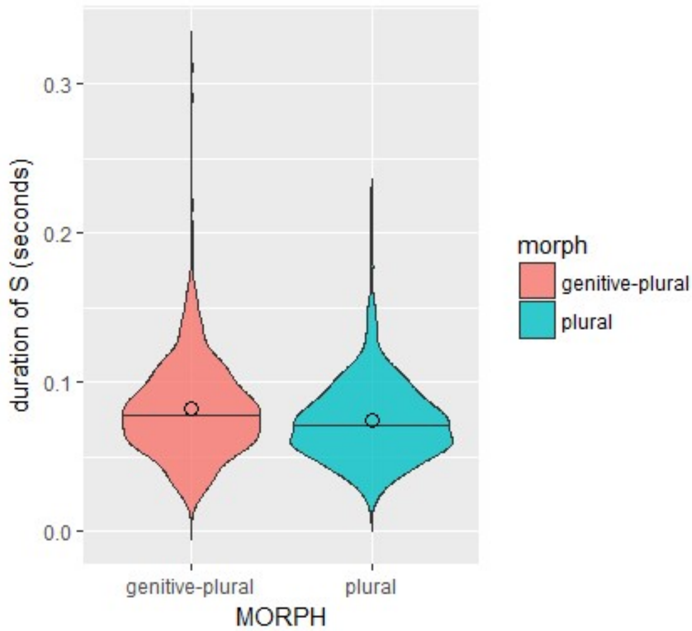


Figure 2: Durations of plural and genitive-plural S. The horizontal line indicates the median, the dot represents the mean.

Before fitting regression models to the data we inspected the distribution of the durations of S. The non-normal distribution with several outliers suggested some trimming or transformation of this variable (see, for example, Baayen & Milín 2010 on issues of data trimming prior to analysis). We tested four different procedures, resulting in four slightly different data sets of slightly differing sample sizes that entered the regression analyses.

Data set 1:
Untransformed dependent variable and exclusion of outliers. We excluded 12 overly long tokens (duration of S>165 ms, N=867).

Data set 2:
Logarithmic transformation and no further data trimming prior to the analysis (N=879).

Data set 3:
Logarithmic transformation plus exclusion of data points smaller or larger than 2.5 standard deviations (N=860)

Data set 4:
Box-Cox transformation (λ=0.14141) and no further trimming prior to the analysis (N=879). The Box–Cox transformation (Box & Cox 1964, Venables & Ripley 2002) is used to identify a suitable transformation parameter λ for a power transformation, and this type of transformation has been implemented successfully in previous studies of affix durations (Plag et al. 2017, Ben Hedia & Plag 2017, Ben Hedia 2019). In the present study the Box-Cox transformation of S durations yielded the same λ (λ=0.14141) in the linear model with MORPH as the variable of interest as in the linear model with LOGWORDFORMFREQ as the variable of interest. This means that we can use data set 4 with both variables of interest.

Eight models were fitted with duration of S as dependent variable, two for each data set. One of these two models contained MORPH as the variable of interest, the other one LOGWORDFORMFREQUENCY. In what follows, the models are numbered according to data sets, and are alphabetically named 'a' or 'b' according to variable of interest ('a' referring to models with MORPH, 'b' to models with LOGWORDFORMFREQUENCY). For instance, 'model 1a' is the model fitted to data set 1 with MORPH as variable of interest, while 'model 2b' is the model fitted to data set 2 with LOGWORDFORMFREQUENCY as variable of interest.

4.2 Results: Duration of S, with MORPH as variable of interest

The models were fitted according to the procedures described above. The final models all contained three significant fixed effects, and either two or three random effects. Table 2 gives an overview of the four final models.

Table 2: Model overview: Transformation of the dependent variable (duration of S), prior trimming, random effect structure and significance levels of the fixed effects in the final models (‘***’<0.001, ‘**’<0.01, ‘*’<0.05).

	Data set 1	**Data set 2**	**Data set 3**	**Data set 4**
model	model 1a	model 2a	model 3a	model 4a
dependent variable	untransformed	logged	logged	Box-Cox
prior trimming	yes	no	yes	no
Random effects				
SUBJECT	✓	✓	✓	✓
ITEM	✓	✓	✓	✓
MORPH BY SUBJECT	✓			✓
Fixed Effects				
MORPH	***	***	***	***
VOICING	***	***	***	***
SPEECHRATE	*	*	*	**

Let us first look at the fixed effects. In all four models the same three variables are significant: MORPH, VOICING and SPEECHRATE. The estimated mean durational difference between plural S and genitive-plural S differs slightly across models. It ranges between 7 and 8 ms for unvoiced S (7.1, 8.0, 7.8, 7.8 ms for models 1a-4a, respectively). The predictive power of the three variables is moderate: The proportion of explained variance that can be assigned to the fixed effects ranges from 12.6 (model 1) to 14.0 percent (model 4) (estimated by using the `r.squaredGLMM` function from the `MuMin` package, Barton 2009).

We assessed the relative importance of the three fixed effects by standardizing these variables by subtracting the mean and dividing it by two standard deviations (see Gelman & Hill 2006: 56f for discussion), and then running the models with these standardized predictors. Table 3 gives the coefficients of the standardized variables in the final models.

Table 3: Effect sizes of MORPH, VOICING and SPEECHRATE, standardized predictors

Fixed effect	**Data set 1**	**Data set 2**	**Data set 3**	**Data set 4**
MORPH.SCALED	-0.0069	-0.092	0.083	0.0090
VOICING.SCALED	-0.0077	-0.109	0.106	0.0127
SPEECHRATE.SCALED	-0.0064	-0.074	0.084	0.0010

In each model, the effect size of VOICING.SCALED is highest, followed by that of MORPH.SCALED. SPEECHRATE.SCALED has the smallest effect size across models. The effect sizes of VOICING.SCALED and MORPH.SCALED are very close to each other in all models. Thismeans that the effect size of voiced vs. unvoiced S is about the same as the one of the difference between plural S and genitive-plural S.

The mixed effect structure is also of interest. The random intercept for item in all models shows that individual words vary in the durations of the S. Furthermore, individual speakers also show differences. In models 1 and 4 the log-likelihood tests showed that a more complex random effect structure was justified, i.e. the inclusion of a random contrast for MORPH by SUBJECT, in addition to the random intercepts for SUBJECT and ITEM. This means that in these data sets there is some evidence that individual speakers may vary in the way the durations of plural and genitive-plural S differ. The inclusion of this random contrast was not justified in models 2 and 3, where the improvement in the AIC did not result in a significantly better fit (model 2: p=0.09, model 3: p=0.41). Models 2 and 3 have in common that the dependent variable is log-transformed, while the other two models have either no transformation or a different one (Box-Cox, with λ=0.14141). It thus seems that speaker variation becomes less pronounced when S durations are log transformed.

We will nevertheless take a closer look at the effect of MORPH by SUBJECT to see how this plays out, and we will do so with data set 1 for illustration (data set 4 behaves very similarly). The estimated contrasts of the 79 individual speakers for the variable MORPH range from -17.0 ms (contrast: from genitive-plural to plural) to 0.6 ms, with a mean of -7.1 ms and a standard deviation of 4.1 ms. Only one speaker has a positive contrast, which is, however, practically zero (0.6 ms), which means that this speaker does not show a difference between the two categories. The distribution of the two categories across speakers is unlikely to be the source of the speaker-dependent variation. There is only a marginally significant correlation between the ratio of plural vs. genitive-plural forms provided by a given speaker and that speaker's random contrast (S=95341, p=0.07, Spearman)

This distribution of the speaker-specific contrasts and the significance of the mean random contrast indicate two things: The average speaker produces a significant durational difference between the two categories, but there is variation, with some speakers showing a very pronounced difference, and other speakers exhibiting a less pronounced difference. Given that log transformation practically eliminates the effect, we refrain from drawing any firm conclusions on speaker variability.

4.3 Results: Duration of S, with LOGWORDFORMFREQ as variable of interest

Models were fitted according to the procedures described in section 3.3. In the three models with transformed S durations the effect of LOGWORDFORMFREQ is very highly significant. In these models, with increasing LOGWORDFORMFREQ the (transformed) duration of S becomes shorter. The effect is illustrated in Figure 3 for model 2b.

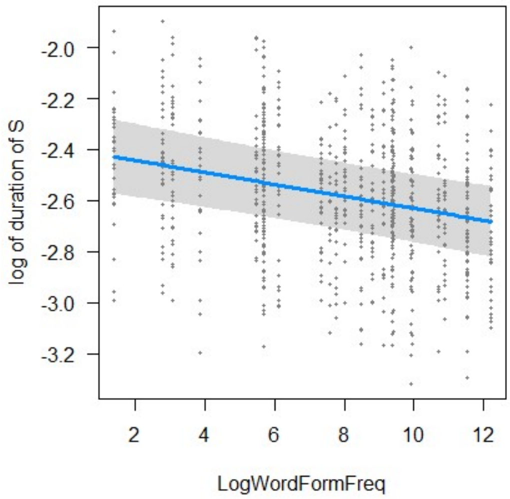


Figure 3: Log-transformed duration of S by log-transformed word-form frequency, as estimated in model 2b

Table 4 summarizes the models. Only in the model with untransformed S durations does LOGWORDFORMFREQ not reach significance (*t*=-1.8, *p*=0.09). We therefore do not give the other effects for this model. With regard to random effects, only the two random intercepts for subject and item justified being included in models 2b, 3b and 4b.

Table 4: Model overview for LOGWORDFORMFREQ as variable of interest: Transformation of the dependent variable, prior trimming, random effect structure and significance levels of the fixed effects in the final models (‘***’<0.001, ‘**’<0.01, ‘*’<0.05).

	Data set 1	**Data set 2**	**Data set 3**	**Data set 4**
model	model 1b	model 2b	model 3b	model 4b
dependent variable	untransformed	logged	logged	Box-Cox
prior trimming	yes	no	yes	no
Random effects				
SUBJECT		✓	✓	✓
ITEM		✓	✓	✓
Fixed Effects				
LOGWORDFORMFREQ	n.s.	***	**	***
VOICING		***	***	***
SPEECHRATE		*	*	***

4.4. Summary of results

Let us summarize our findings concerning the hypotheses we set up in section 2. We repeat them here for convenience.

(2’) H0 ‘Structural Hypothesis’
There is no difference in duration between the plural S and the genitive-plural S.

(3’) H1 ‘Spelling Hypothesis’
Plural S is shorter than genitive-plural S

(4’) H2 ‘Prosody Hypothesis’
Plural S is longer than genitive-plural S

(5’) H3 ‘Morpheme-based Frequency Hypothesis’
Plural S is shorter than genitive-plural S

(6’) H4 ‘Complexity Hypothesis’
Plural S is shorter than genitive-plural S

(7’) H5 ‘Word-based Frequency Hypothesis’
a. The duration of final S depends on the frequency of the word-form it is part of. The more frequent the word-form, the shorter the S.

H0 has to be rejected, there are significant differences in duration between plural S and genitive S in all models, with the genitive-plural S being longer by about 7 or 8 ms. This difference is in accordance with H1, H3 and H4. H2 has to be rejected, the difference goes in the opposite direction of the one predicted by H2. Finally, H5a is also supported by our data.

We will discuss the implications of these results after having looked at the influence of our variables of interest on word duration.

5. Results: Word duration

5.1 Data sets

The distribution of this dependent variable made some trimming necessary for untransformed word durations, log-transformed word durations and Box-Cox-transformed word durations. We created four data sets:

Data set 5:
Untransformed word durations as the dependent variable. 16 outliers with durations of more than 870 milliseconds or durations of less than 210 milliseconds were removed after manual inspection of the distribution. This resulted in N=863.

Data set 6:
Log-transformation of word durations; removal of items with standardized values that are smaller than -2.5, or larger than 2.5 standard deviations. This resulted in N=869.

Data set 7:
Box-Cox-transformation of word durations, based on a linear model with MORPH as the variable of interest (λ=-0.1818182); removal of items with standardized values that are smaller than -2.5, or larger than 2.5 standard deviations (N=867).

Data set 8:
Box-Cox-transformation of word durations, based on a linear model with LOGWORDFORMFREQ as the variable of interest (λ=-0.1818182); removal of items with standardized values that are smaller than -2.5, or larger than 2.5 standard deviations (N=867).

5.2 Results: Word duration

We fitted the data sets according to the procedures described in section 3.3. We also included analyses with MORPH as variable of interest in order to see if the morphological categories have an influence beyond the final S. Table 5 gives an overview of the significant effects in the final models.

Table 5: Overview of regression models with word duration as dependent variable (Significance levels: '***'<0.001, '**'<0.01, '*'<0.05).

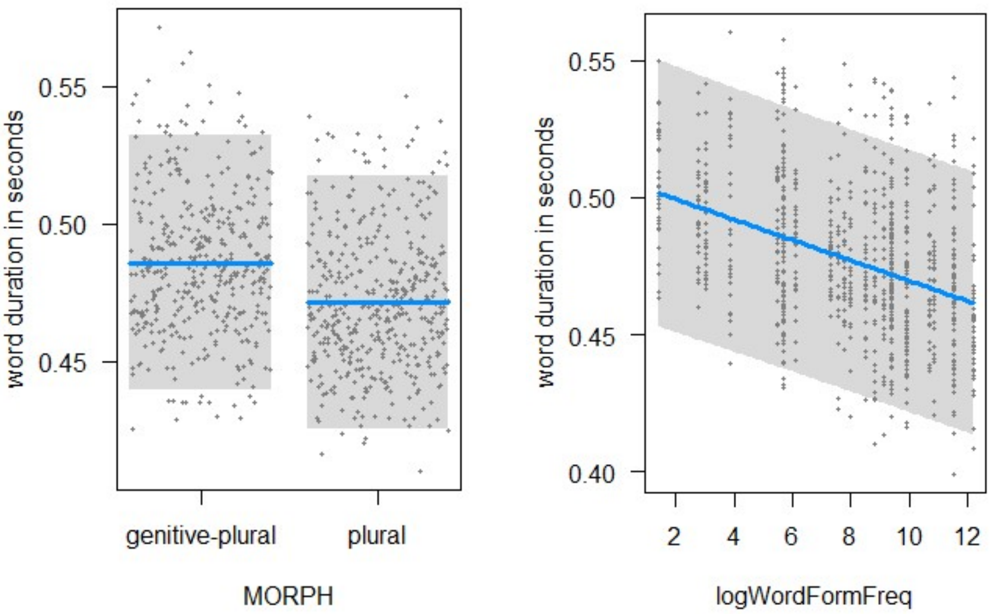
	Data set 5	**Data set 6**	**Data set 7**
model	model 5a	model 6a	model 7a
variable of interest		**MORPH**	
dependent variable	untransformed	logged	Box-Cox
Random effects			
SUBJECT	✓	✓	✓
ITEM	✓	✓	✓
MORPH BY SUBJECT	✓	✓	✓
Fixed Effects			
MORPH	***	***	***
VOICING	**	***	***
SPEECHRATE	***	***	***
NUMBEROFSYLLABLES	***	***	***

	Data set 5	**Data set 6**	**Data set 8**
model	model 5b	model 6b	model 8
variable of interest	**LOGWORDFORMFREQ**		
dependent variable	untransformed	logged	Box-Cox
Random effects			
SUBJECT	✓	✓	✓
ITEM	✓	✓	✓
LOGWORDFORMFREQ BY SUBJECT	✓		✓
Fixed Effects			
LOGWORDFORMFREQ	***	***	***
VOICING	***	***	***
SPEECHRATE	***	***	***
NUMBEROFSYLLABLES	***	***	***

The random effect structures vary across models, but all models apart from model 6b show a significant effect of random contrasts/slopes by SUBJECT for the variable of interest, i.e. MOPRH or LOGWORDFORMFREQ. Speakers thus vary in how susceptible they are with regard to strength of the effect of MORPH or LOGWORDFORMFREQ. The effects of the significant covariates are as expected and in line the literature.

All data sets show an effect of the respective variable of interest. Both morphological category and word-form frequency are significantly related to word duration. On average, plural words have shorter durations than genitive-plural words by about 14 ms, which means that it is not only the suffix (which showed a difference of about 7 ms) that is affected by this factor. With regard to word-form frequency we can say that with rising word-form frequency, word duration decreases. Figure 4 illustrates these two effects for data set 5, i.e. the data set with the untransformed word durations.

Figure 4: Left panel: Word duration by MORPH, data set 5, model 5a. Right panel: Word duration by LOGWORDFORMFREQ, data set 5, model 5b.



These results relate to the hypotheses in a similar fashion as the results on S duration. H0 has to be rejected, there are significant differences in word duration between plural S and genitive S in all models, with the genitive-plural word-forms being longer by about 14 ms. This difference is in accordance with H1, H3 and H4. H2 has to be rejected, too, it goes in the wrong direction.

H5b stated that the duration of the whole word, and not only the duration of the final S, depends on the frequency of the word-form. This turned out to be true. The difference in word duration is about double the difference we found for final S. That is, the effect of word-form frequency of duration is distributed with about one half over the stem and with the other half over the final S.

6. Discussion and conclusion

To summarize, the statistical analysis has shown that plural S and genitive-plural S differ significantly in duration, with the genitive-plural S being longer by about 7 or 8 ms. This refutes the null hypothesis (H0) and the Prosody Hypothesis (H2). This means that a purely structural morphological approach cannot account for the durational patterns, nor can an account that wants to relate the durational differences to prosodic structure.

The Spelling Hypothesis (H1) is supported by our data. Previous studies on the relationship of orthography and acoustic duration have also found that the number of orthographic symbols representing a sound correlates with the duration of that sound in speech. Unfortunately, there is no theory available that can account for this patterning of spelling and acoustic duration. It is presently still unclear how orthographic effects on speech can be generally accounted for in a realistic model of articulation or speech production. With regard to <s’> there might be the additional complication that the apostrophe has as one of its conventionalized functions that it replaces something that is missing. Viewed from this angle the use of the apostrophe mirrors the idea that there are two S’s at some level of representation. Morphology and spelling are therefore inextricably linked when it comes to the spelling of plural and genitive-plural, which makes it hard to tease apart potential morphological effects and potential orthographic effects.

The Complexity Hypothesis (H4) is also supported by the data. This result would seem to support theories in which the processing of the more complex morpho-syntactic feature specification (i.e. the genitive-plural) slows down production of the exponent of that feature specification. However, the increased morpho-syntactic complexity of the genitive-plural coincides with its lower frequency, so that the Morpheme-based Frequency Hypothesis (H3) is also confirmed by our study.

As was shown in our analyses of the word durations, the durational difference between plurals and genitive-plurals is not restricted to the final S. The duration of the whole word-form varies by morphological category. Plural words are 14 ms shorter than genitive-plural words, on average. Again, there is no theory that can straightforwardly account for this difference. If we, however, include word-form frequency into our considerations, we might find an answer why genitive-plurals are longer than plurals.

Both Word-based Frequency Hypotheses (H5a and H5b) were confirmed by our data. Higher word-form frequencies go together with shorter S durations and with shorter word durations. How does this finding relate to the effects of morpho-syntactic category discussed so far? Overall we can state that the word-form frequency effect on duration holds across the board, i.e. all word-forms are affected by it, irrespective of the morphological category expressed. First, this means that the plural forms of two different lexemes show a durational difference provided that the two forms have sufficiently different word-form frequencies. For instance, in our data set, the plural of *boy* has a log word-form frequency of 10.7, while the plural form of *dog* has one of 9.9. The word durations pattern as expected: *boys* has a mean duration of 311 ms, while *dogs* (the less frequent form) is 413 ms long on average.

Second, we saw that genitive-plural word-forms are all less frequent than their corresponding plural word forms (as can be seen in the right panel of Figure 1). It is therefore reasonable to assume that the significant difference in the mean word duration between plural nouns and genitive-plural nouns arises from the fact that the average word-form frequency of the plural is much higher than that of the genitive-plural. This line of reasoning is corroborated by a look at the pairwise distribution of word durations, as shown in figure 5. Like in Figure 1, each pair of dots represents one lexeme with its plural and the genitive-plural forms, respectively.

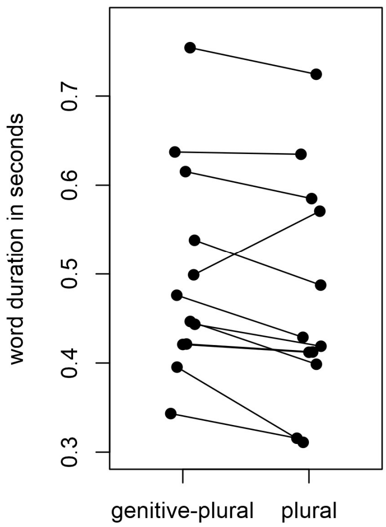


Figure 1: Word duration by morphological category

We see that for all pairs but one the genitive-plural form is longer than the corresponding plural form (only the lexeme *patient* shows the opposite behavior). This means that the effect of morphological category on duration can be attributed to an underlying effect of word-form frequency: the plural forms are shorter because they are more frequent. To further substantiate this conclusion we carried out additional analyses with the frequency ratio of plural and plural-genitive as variable of interest (following Lohmann's 2018 analogous analysis of homophonous lexemes such as *time* and *thyme*). This ratio captures the difference in frequency between a given plural word-form, e.g. *dogs*, compared to the corresponding genitive-plural word-form, e.g. *dogs'*. All data sets showed a significant effect of the frequency ratio on the duration of S and the duration of the whole word in the expected direction. The larger the difference in word-form frequency between the corresponding plural and genitive-plural forms of a lexeme, the larger the difference between the (shorter) duration of the plural, and the (longer) duration of the genitive-plural. The models are included in Appendix B (see the columns with LOGWORDFORMFREQRATIO as the variable of interest).

In sum, word-form frequency is predictive of duration (across and within morphological categories), resulting in an average difference in duration between plurals and genitive-plurals. This is in line with other studies of the production of inflected words (Caselli et al. 2015, Lõo et al. 2018) that have demonstrated that less frequent word-forms are pronounced with longer duration.

Our results also have implications for morphological theory. Word-form frequency effects for regularly inflected words in speech production are at odds with theories in which storage only plays a role for morphologically irregular words, or for highly frequent regular words (e.g. Pinker 1999, Alegre & Gordon 1999). Our data include very rare word-forms, but the frequency effect is nevertheless observable with these forms.

The word-form frequency effect could be more naturally accounted for in word-and-paradigm models of morphology (e.g. Matthews 1974, Blevins 2016), in which individual word-forms may have representations in a network of morphologically related forms. In a more modern perspective on word-and-paradigm organization, word-and-paradigm effects may also arise without static representations in the mental lexicon, but by dynamic states of the cognitive system that are constantly updated on the basis of new input (Tomaschek et al. 2019, Baayen et al. 2019).

To summarize, this article has shown that, phonetically, plurals and corresponding genitive plurals in English are not homophonous. The fact that complex words may display interesting durational characteristics depending on their morphological makeup has implications for our thinking about lexical organization and lexical processing, and we hope to have shown that the analysis of fine phonetic detail of complex words can inform both speech production models and morphological theory.

Acknowledgements

This study is part of an ongoing collaboration within the DFG Research Unit FOR2373 'Spoken Morphology'. We are very grateful to the Deutsche Forschungsgemeinschaft for funding this research (Grants: LO-2135/1-1 'The Phonetics of Word Class and its Representation in the Lexicon' to Arne Lohmann; PL151/8-1 and PL151/8-2 'Morpho-phonetic Variation in

English' to Ingo Plag; PL151/7-1 and PL151/7-2 'FOR 2737 Spoken Morphology: Central Project' to Ingo Plag).

References

Alegre, M. & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40(1), 41–61.

Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics.* Cambridge: Cambridge University Press.

Baayen, R. H., Chuang, 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*, 2019(1), 1-39.

Baayen, R. H. & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28.

Barton, K. (2009). MuMIn: Multi-model inference. Software package. http://r-forge.r-project.org/projects/mumin.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. http://CRAN.R-project.org/package=lme4.

Bauer, L., Lieber, R. & Plag, I. (2013). *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Ben Hedia, S. (2019). *Gemination and degemination in English affixation: Investigating the interplay between morphology, phonology and phonetics.* Berlin: Language Science Press.

Ben Hedia, S. & Plag, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics*, 62, 34–49.

Blazej, L. J. & Cohen-Goldberg, A. M. (2015). Can we hear morphological complexity before words are complex? *Journal of Experimental Psychology: Human Perception and Performance*, 41(1), 50–68.

Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford: Oxford University Press.

Boersma, P. & Weenink, D. (2016). *Praat: doing phonetics by computer*. [Computer program]. Version 6.0.14. http://www.praat.org/.

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society*, B 26, 211–252.

Brewer, J. (2008). *Phonetic reflexes of orthographic characteristics in lexical representation*. Tucson: University of Arizona PhD dissertation.

Caselli, N. K., Caselli, M. K. & Cohen-Goldberg, A. M. (2016). Inflected words in production: Evidence for a morphologically rich lexicon. *Quarterly Journal of Experimental Psychology*, 69(3), 432-454.

Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement. *Morphology*, 24(4), 291–323.

Davies, M. (2013). *The Corpus of Contemporary American English (full text on CD): 440 million words*, 1990-2012.

Gelman, A. & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

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

Kemps, R., Ernestus, M., Schreuder, R. & Baayen, R. H. (2005). Prosodic cues for morphological complexity: The case of Dutch noun plurals. *Memory and Cognition*, 33(3), 430–446.

Kisler, T., Reichel, U. & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326–347.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59(5), 1208–1221.

Kunter, G. (2016). *Coquery: a free corpus query tool*. http://www.coquery.org.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1-26.

Levelt, W. J., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38.

Lohmann, A. 2018. Time and thyme are not homophones: A closer look at Gahl's work on the lemma-frequency effect, including a reanalysis. *Language* 94(2), e180-e190.

Lohmann, A. & Conwell, E. (2019). Phonetic effects of grammatical category: How category-specific prosody and token frequency impact the acoustic realization of nouns and verbs. *Submitted*.

Lõo, K., Järvikivi, J., Tomaschek, F., Tucker, B. V. & Baayen, R. H. (2018). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects.

Morphology, 28(1), 71–97.

Matthews, P. H. (1974). *Morphology. An Introduction to the Theory of Word Structure*. London: Cambridge University Press.

Palmer, F., Huddleston, R. & Pullum, G. K. (2002). Inflectional morphology and related matters. In R. D. Huddleston & G. K. Pullum, eds., *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, pp. 1565–1619.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language.* London: Weidenfeld and Nicolson.

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

Plag, I. (1998). Morphological haplology in a constraint-based morpho-phonology. In W. Kehrein & R. Wiese, eds., *Phonology and Morphology of the Germanic Languages*. Tübingen: Niemeyer, pp. 199–215.

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

Rose, D. (2017). *Predicting plurality: An examination of the effects of morphological predictability on the learning and realization of bound morphemes.* Christchurch: University of Canterbury PhD Dissertation.

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F. & Malouf, R. (2018). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience*, 33(1), 32–49.

Tomaschek, F., Hendrix, P. & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, 71, 249–267.

Tomaschek, F., Plag, I., Ernestus, M. & Baayen, R.H. (2019). Modeling the duration of word-final S in English with Naive Discriminative Learning. To appear in *Journal of Linguistics*.

Tomaschek, F., Tucker, B. & Baayen, R. H. (submitted). How is anticipatory coarticulation of suffixes affected by linguistic proficiency?

Turk, A. & Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 1–13.

van Borsel, J. & De Maesschalck, D. (2008). Speech rate in males, females, and male-to-female transsexuals. *Clinical Linguistics & Phonetics*, 22(9), 679–685.

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S-Plus*, 4th edn., New York: Springer.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, 91(3), 1707–1717.

Zimmermann, J. (2016). Morphological status and acoustic realization: Findings from New Zealand English. In C. Carignan & M. D. Tyler, eds., *Proceedings of the 16th Australasian International Conference on Speech Science and Technology*. Sydney: University of Western Sydney, pp 6-9.

Zimmermann, J., Rose, D., Bürkle, D. & Watson, K. (2017). The Role of Predictability and Sub-Phonemic Detail in Speech Perception: English has-Clitic [s] vs. Plural [s]. Paper presented at *Old World Conference on Phonology* 2017, 20–22 February 2017, Düsseldorf.

Zwicky, A. M. (1975). Settling on an underlying form: The English inflectional endings. In D. Cohen & J. R. Wirth, eds., *Testing linguistic hypotheses*. New York: Wiley, pp. 129–185.

Appendix A

Paragraphs and sentences used in the present study, from Lohmann & Conwell's experiment. The material in parentheses represents the sentence version with long/extended PP.

Ben and Susan wonder why their teacher always gets aggravated at the theater. They realize it's because of the chaperones' bad behavior.
Noun sentence: The parents' chat during the play (on US history) angers Mr. Robinson.
Verb sentence: The parents chat during the play (on US history) and this angers Mr. Robinson.

Ms. Butler, the science teacher, comments to her colleague that her students are very talkative before exams. She suggests that there is a reason for this.
Noun sentence: The students' chat about the quiz (on advanced Chemistry) makes them feel more confident.
Verb sentence: The students chat about the quiz (on advanced Chemistry) and this makes them feel more confident.

When the children visit their relatives, everything is different. They never know what to expect.
Noun sentence: Their grandparents' cook with the bright clothing (from India) entertains Louis and Robin.
Verb sentence: Their grandparents cook with special spices (from India) and this delights Louis and Robin.

The Hendersons were known to be wealthy and flamboyant. They hosted a large party following the annual travel agents' meeting.
Noun sentence: The Hendersons' cook for the reception (at the conference) entertains the invited guests.
Verb sentence: The Hendersons cook for the reception (at the conference) and this delights the invited guests.

Maria and Pedro had their property landscaped by a garden designer. One day, their neighbor's dogs come through a hole in the fence.
Noun sentence: The dogs' dig behind the shed (in the yard) upset Maria and Pedro.
Verb sentence: The dogs dig behind the shed (in the yard) and this upsets Maria and Pedro.

Natalie and Carson are pretending to be archaeologists. They put on pith helmets and took their shovels across the street.

Noun sentence: The kids’ dig at the playground (in the park) entertains the parents.

Verb sentence: The kids dig at the playground (in the park) and this entertains the parents.

The gossip surrounding the famous couple has been building for weeks. Everyone who interacts with them is getting really tired of it.

Noun sentence: The actors’ kiss on the movie set (for the new production) annoys the director.

Verb sentence: The actors kiss on the movie set (for the new production) and this annoys the director.

At the premiere of the new play Steve manages to sneak behind the stage. From his spot in the corner he witnesses an argument between the director and some of the actors.

Noun sentence: The actors’ look through the curtains (of the theater) irritates the director.

Verb sentence: The actors look through the curtains (of the theater) and this irritates the director.

Mike and his team are very busy finishing up the report for the end of the quarter. They see that some of their co-workers in accounting do not seem to take their work seriously.

Noun sentence: Their colleagues’ nap in the cubicle (next to the busy hallway) upsets the hard-working employees.

Verb sentence: Their colleagues nap in the cubicle (next to the busy hallway) and this upsets the hard-working employees.

Dr. Butler and Dr. Gonzales have moved their practice out of the city. Now, some of the older patients are very sleepy when they arrive at the cardiologists’ new office.

Noun sentence: The patients’ nap in the waiting room (with the new furniture) irritates the doctors.

Verb sentence: The patients nap in the waiting room (with the new furniture) and this irritates the doctors.

Peter and JJ were playing by the school when some dark clouds rolled in. Their mother had told them to keep their things inside in case of rain, but they didn’t listen.

Noun sentence: The young boys' pack under the tree (near the playground) got wet in the rain.

Verb sentence: The young boys pack under the tree (near the playground) and get wet in the rain.

After they found their cabin, Barb and Todd began getting ready for the next day. They wanted to get an early start, so Todd got everything organized.

Noun sentence: The hikers' pack for the long hike (in the mountains) was prepared the night before.

Verb sentence: The hikers pack for the long hike (in the mountains) and prepare the night before.

Corporations aren't always concerned with what's best for the Earth. When oil prices are high, they stop at nothing to extract more and more.

Noun sentence: The oil corporations' push for extensive investment (in the fracking sector) worries environmentalist groups.

Verb sentence: The oil corporations push for extensive investment (in the frackingsector) and this worries environmentalist groups.

Appendix B

Table B.1: Regression models with duration of S as dependent variable. For the fixed effects, the table gives the coefficients, standard errors are given in parentheses. Significance codes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

	Model 1a	Model 2a	Model 3a	Model 4a	Model 1b	Model 2b	Model 3b	Model 4b	Model 1c	Model 2c	Model 3c	Model 4c
(Intercept)	0.0867***	-2.4289***	-2.4387***	0.7124***	0.0829***	-2.2883***	-2.3071***	0.7268***	0.0836***	-2.4746***	-2.4275***	0.7098***
	(0.0050)	(0.0723)	(0.0701)	(0.0070)	(0.0049)	(0.0885)	(0.0935)	(0.0086)	(0.0049)	(0.0718)	(0.0708)	(0.0072)
MORPHplural	-0.0071***	-0.0946***	-0.0936***	-0.0090***								
	(0.0015)	(0.0183)	(0.0176)	(0.0021)								
VOICINGvoiced	-0.0083***	-0.1411***	-0.1486***	-0.0128***	-0.0096***	-0.1422***	-0.1160***	-0.0135***	-0.0083***	-0.1454***	-0.1318***	-0.0129***
	(0.0017)	(0.0239)	(0.0229)	(0.0023)	(0.0020)	(0.0238)	(0.0308)	(0.0023)	(0.0016)	(0.0237)	(0.0243)	(0.0025)
SPEECHRATE	-0.0005*	-0.0081*	-0.0070*	-0.0010**	-0.0008**	-0.0085*	-0.0106*	-0.0011***	-0.0006*	-0.0081*	-0.0120***	-0.0012***
	(0.0002)	(0.0035)	(0.0034)	(0.0003)	(0.0003)	(0.0035)	(0.0041)	(0.0003)	(0.0002)	(0.0035)	(0.0034)	(0.0004)
LOGWORDFORMFREQ						-0.0235***	-0.0200**	-0.0022***				
						(0.0050)	(0.0063)	(0.0005)				
WORDFORMFREQRATIO									-0.0024***	-0.0318***	-0.0300***	-0.0025***
									(0.0007)	(0.0059)	(0.0060)	(0.0006)
AIC	-4087.8488	241.7642	164.2789	-3604.5214	-3977.5509	233.1821	658.0965	-3597.6545	-4090.2107	229.6394	289.8290	-3551.6202
BIC	-4045.4870	274.8226	197.2098	-3562.0178	-3929.9005	266.2152	691.1969	-3564.5960	-4038.4352	262.6809	322.9462	-3518.3706
Log Likelihood	2052.9244	-113.8821	-75.1394	1811.2607	1998.7754	-109.5910	-322.0483	1805.8272	2056.1053	-107.8197	-137.9145	1782.8101
Num. obs.	818	831	816	831	867	828	836	831	818	829	838	854
Num. groups: SUBJECT	78	78	78	78	78	78	78	78	78	78	78	78
Num. groups: ITEM	12	12	12	12	12	12	12	12	12	12	12	12
Var: SUBJECT (Intercept)	0.0001	0.0066	0.0065	0.0001	0.0001	0.0062	0.0092	0.0001	0.0000	0.0062	0.0057	0.0001
Var: SUBJECT MORPHplural	0.0000			0.0001	0.0001							

Cov: SUBJECT (Intercept) MORPHplural	-0.0000			-0.0001	-0.0001							
Var: ITEM (Intercept)	0.0002	0.0338	0.0324	0.0003	0.0003	0.0518	0.0405	0.0005	0.0002	0.0335	0.0333	0.0003
Var: Residual	0.0003	0.0668	0.0606	0.0006	0.0005	0.0657	0.1109	0.0006	0.0003	0.0658	0.0711	0.0008
Var: ITEM MORPHplural					0.0001							
Cov: ITEM (Intercept) MORPHplural					-0.0001							
Var: SUBJECT WORDFORMFREQRATIO									0.0000			
Cov: SUBJECT (Intercept) WORDFORMFREQRATIO									-0.0000			
Var: ITEM WORDFORMFREQRATIO									0.0000			
Cov: ITEM (Intercept) WORDFORMFREQRATIO									-0.0000			

Table B.2: Regression models with word duration as dependent variable. For the fixed effects, the table gives the coefficients, standard errors are given in parentheses. Significance codes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

	Model 5a	**Model 6a**	**Model 7a**	**Model 5b**	**Model 6b**	**Model 8a**	**Model 5c**	**Model 6c**	**Model 9c**
(Intercept)	0.4555^{***}	-0.7616^{***}	1.1476^{***}	0.4807^{***}	-0.7050^{***}	1.1373^{***}	0.4471^{***}	-0.7733^{***}	1.1503^{***}
	(0.0615)	(0.1528)	(0.0329)	(0.0638)	(0.1590)	(0.0339)	(0.0614)	(0.1527)	(0.0329)
MORPHplural	-0.0144^{***}	-0.0240^{***}	0.0048^{***}						
	(0.0029)	(0.0053)	(0.0011)						
VOICINGvoiced	-0.0081^{**}	-0.0218^{***}	0.0044^{***}	-0.0100^{***}	-0.0198^{***}	0.0046^{***}	-0.0093^{***}	-0.0228^{***}	0.0046^{***}
	(0.0025)	(0.0051)	(0.0011)	(0.0025)	(0.0052)	(0.0011)	(0.0025)	(0.0051)	(0.0011)
SPEECHRATE	-0.0242^{***}	-0.0566^{***}	0.0120^{***}	-0.0236^{***}	-0.0574^{***}	0.0120^{***}	-0.0241^{***}	-0.0566^{***}	0.0120^{***}
	(0.0005)	(0.0009)	(0.0002)	(0.0005)	(0.0010)	(0.0002)	(0.0005)	(0.0009)	(0.0002)
NUMBEROFSYLLABLES	0.1695^{***}	0.3637^{***}	-0.0761^{***}	0.1637^{***}	0.3581^{***}	-0.0747^{***}	0.1694^{***}	0.3634^{***}	-0.0763^{***}
	(0.0272)	(0.0677)	(0.0146)	(0.0281)	(0.0703)	(0.0150)	(0.0272)	(0.0677)	(0.0146)
LOGWORDFORMFREQ				-0.0037^{***}	-0.0062^{***}	0.0012^{***}			
				(0.0006)	(0.0011)	(0.0002)			
WORDFORMFREQRATIO							-0.0046^{***}	-0.0076^{***}	0.0015^{***}
							(0.0009)	(0.0017)	(0.0003)
AIC	-3388.8972	-2265.9568	-4840.8887	-3408.0084	-2248.4654	-4825.1675	-3400.8426	-2263.7663	-4839.2058
BIC	-3341.6950	-2218.7185	-4793.6986	-3360.8062	-2210.6652	-4777.9412	-3353.6646	-2216.5279	-4792.0156
Log Likelihood	1704.4486	1142.9784	2430.4443	1714.0042	1132.2327	2422.5838	1710.4213	1141.8831	2429.6029
Num. obs.	829	832	828	829	833	831	827	832	828
Num. groups: SUBJECT	78	78	78	78	78	78	78	78	78
Num. groups: ITEM	12	12	12	12	12	12	12	12	12
Var: SUBJECT (Intercept)	0.0004	0.0010	0.0000	0.0010	0.0003	0.0000	0.0001	0.0003	0.0000
Var: SUBJECT MORPHplural	0.0004	0.0010	0.0000						
Cov: SUBJECT (Intercept) MORPHplural	-0.0004	-0.0010	-0.0000						
Var: ITEM (Intercept)	0.0066	0.0407	0.0019	0.0070	0.0439	0.0020	0.0066	0.0407	0.0019
Var: Residual	0.0007	0.0029	0.0001	0.0007	0.0032	0.0001	0.0007	0.0029	0.0001
Var: SUBJECT LOGWORDFORMFREQ				0.0000		0.0000			
Cov: SUBJECT (Intercept) LOGWORDFORMFREQ				-0.0001		-0.0000			
Var: SUBJECT WORDFORMFREQRATIO							0.0000	0.0001	0.0000
Cov: SUBJECT (Intercept) WORDFORMFREQRATIO							-0.0000	-0.0001	-0.0000