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 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 as paradigmatic probability (Cohen 2014).

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). In addition, Seyfarth et al. (2017) and Engemann et al. (2019) found that the duration of the portion of the word that precedes final S varies as a function of the kind of S (plural vs. non-morphemic). In both studies morphemic final S goes together with longer duration of the preceding material. For some types of S-final words, especially for genitive-plurals, 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 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). The paper has two main aims. First, we want to explore whether plurals and genitive-plurals show differences in acoustic duration. The results of this investigation have important theoretical implications, as they can be interpreted against the backdrop of competing the predictions made by two kinds of approaches concerning the durational properties of plural words and genitive-plural words in English. The second aim of this paper is therefore to test these predictions.

On the one hand we test the predictions made by traditional structuralist linguistic theories as well as modular theories of phonology-morphology interaction (e.g. Lexical Phonology) and of speech production (e.g. Levelt, Roelofs & Meyer 1999). Henceforth we will refer to these approaches as ‘structuralist-modular’. These structuralist-modular approaches share the assumption that the phonetic realization of complex words is phoneme-based and that there is no direct interface between morphology and phonetics. According to these approaches, there should be no difference in the phonetic realization of homophonous plural and genitive-plural forms.

On the other hand, we test the predictions made by interactive models of speech production (e.g. Dell 1986). In contrast to the structuralist-modular approaches, these approaches would allow for cascading activation across different levels in the mental lexicon (semantics, phonology, morphology), to the effect that the strength of lexical activation may also influence articulation, and thus the phonetic properties of words. For example, it has been shown that the duration of words in speech is influenced by their lemma frequency (e.g. Gahl 2008, Lohmann 2018) and by the frequency of the inflected word form representing a lemma (henceforth ‘word-form frequency’) (Caselli et al. 2015, Lõo et al. 2017). The exact mechanisms according to which these effects emerge in speech production are not entirely clear, but, at a general level, it seems that articulation is influenced by the ease with which all kinds of lexical information, including morphological information, is processed (e.g. Bell et al. 2009).

The theoretical debate about the role of fine phonetic detail in the production of complex words has important implications also for general theories of morphological organization, in particular the question of morpheme-based vs. word-based approaches to word structure. Word-form frequency effects are not easily accommodated in morpheme-based morphological theories, but quite expected in word-and-paradigm approaches.

In this paper, we report the results of an analysis of data that were elicited in an experiment in which sentences were read aloud which contained plural words and genitive-plural words in very similar contexts (see Lohmann & Conwell 2019, who provided the original data set). 462 plural tokens and 417 genitive-plural tokens were phonetically annotated, and the duration of S as well as the duration of the whole word were 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 on average than genitive-singular nouns. We argue that these results are incompatible with structuralist-modular approaches to morpho-phonology and speech production, and support non-modular theories that allow for the possibility that lexical properties influence phonetic detail.

2. Plural and genitive-plural in English

Traditionally, phonetics plays no role in morphological theorizing. While it is generally acknowledged that morphological structure may interact with phonological structure, morphology is not thought to directly influence phonetic detail. Standard descriptions of morpho-phonological phenomena do not refer to phonetic realizations, or sometimes even explicitly deny the relevance of such considerations. Take, for instance, a classic case of allomorphy. The standard literature (e.g. Palmer et al. 2002, Bauer et al. 2013: chapter 1) holds that there are three different allomorphs of the regular plural S: /ɪz/ after sibilants, /z/ after voiced sounds, /s/ after unvoiced sounds. According to Bauer et al. (2013: 15) "[t]his allomorphy is easily understood in phonological [sic] terms (assimilation and epenthesis to break up illegal geminates), and is not controversial". Bauer et al. do not mention phonetics at all.

The regular genitive-plural has exactly the same allomorphs as the plural, 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). Palmer et al. mention the phonetic level, compare 'in speech', but assume that there is nothing at this level that might distinguish plural and genitive-plural forms. Notice, however, that in writing, the genitive feature is represented by an apostrophe following the plural <s> (as in *boys', dogs'*, Bauer et al. 2013:144f).[1]

---

[1] 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

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.

How may the durations of plural and genitive-plural differ from each other? Different kinds of approaches make conflicting predictions concerning the behavior of the two categories. We will discuss them in turn, starting with two structuralist approaches.

2.1 Structuralist-modular approaches

The first structuralist approach we will discuss is what we might 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 phonological rules or constraints as are used for the selection of the right 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, including the exponents for the feature specification <genitive,

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.

singular> (which has an additional exponent, ø, which may occur with proper nouns, e.g. *Burns'*, or in set expressions, e.g. *for goodness' sake*). 'X' stands for the phonological specification of the base.

(1)	feature specification	exponents morpheme-based	exponents word-based
	<plural>	/z/, /s/ or /ɪz/	/Xz/, /Xs/ or /Xɪz/
	<genitive, singular>	ø, /z/, /s/ or /ɪz/	/X/ø, /Xz/, /Xs/ or /Xɪz/
	<genitive, plural>	/z/, /s/ or /ɪz/	/Xz/, /Xs/ or /Xɪz/

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

The second structuralist 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.).[2] The end result of this would be the same as in the selection approach, the presence of only one segment, with the result that there no phonetic difference between plural and plural-genitive is predicted.

To summarize, a phoneme-based structuralist theory predicts that there be no difference in duration between the plural S and the genitive-plural S. The same prediction emerges from Lexical Phonology (e.g. Kiparsky 1982, Bermúdez-Otero 2017) and modular models of speech production (e.g. Levelt, Roelofs & Meyer 1999). In Lexical Phonology the only formal representation of morphemes is phonological in nature, and phonetic detail is delegated to 'post-lexical' processes, which are taken to be insensitive to the morphological structure of the word. The mechanism of 'bracket erasure' (e.g. Kiparsky 1982) encapsulates this idea. Bracket erasure means that morphological boundaries are no longer visible on the next derivational cycle within the lexicon, nor after the item's emergence from the lexicon.

[2] 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.

Modular feed-forward speech production models (e.g. Levelt, Roelofs & Meyer 1999) also hold that morphological structure is no longer accessible at the level of phonetic encoding. In these models, lexemes are stored in the mental lexicon with their meanings and their phonological representations. These phonological representations are the input for the module called ‘articulator’, which, crucially, does not have access to information about the lexical origin of a sound. A given string of syllabified phonemes in a given context will therefore always be articulated in the same way, irrespective of its morphemic status, and only modulo the variation originating from purely phonetic sources such as speech rate, context, or prosodic phrasing.

2.2 Interactive models of speech production

In contrast to modular models of speech production, there are interactive models of speech production in which lexical activation can spread in less restricted ways (e.g. Dell 1986). Such models allow for the spreading of activation across representations of different kinds, thus enabling mutual influence of entities that would be strictly separated in modular models. For example, within this class models, the morphological or lexical level may yield a more direct influence on the phonological and phonetic levels, as the different entities on these levels do not belong to strictly separated modules. In such an interactive architecture the strength of activation becomes an important determinant for various behavioral measurements, such as reaction times in speech perception, eye-movements in reading, or durations in speech production. In such a model, it is thus possible that two categories, such as plural and genitive plural, display differences in their acoustic realization because of differences in their lexical activation. Frequency or probability of occurrence is taken to be an important correlate of lexical activation, and indeed it has been shown that frequency may significantly influence phonetic durations. It is well-known that words are phonetically reduced, i.e. pronounced shorter, with increasing frequency (see Jurafsky et al. 2001, Gahl 2008), or with increased contextual probability. In our case, the two categories in question are characterized by a considerable difference in usage frequency, with the plural outnumbering the genitive-plural by far.

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 plurals are shorter than genitive-plurals. With regard to word-form frequency effects, Caselli et al. (2015) found that word-form frequency predicts the duration of English words suffixed with *-ing*, *-ed*, and *-s*. 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.

Another explanation for durational differences is processing complexity. According to this idea, speakers may slow down because they have to compute more complex morphology (see Seyfarth et al. 2017: 12f for discussion). With regard to plural vs. genitive-plural, 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. This may result in a longer duration of the genitive-plural S (assuming a morpheme-based view), or in a longer duration of the genitive-plural word-form.

2.3 Summary

To summarize, there are two main approaches that make predictions about a potential difference between plurals and genitive plurals. Under the structuralist-modular approach one does not expect a difference in duration between the two categories. Under the interactive approach a difference in duration between plurals and genitive-plurals might emerge due to differences in their frequencies. or a difference in the number of morphological features that need to be activated in production. The nature of the potential frequency effect will also shed light on the discussion of morpheme- vs. word-based approached to morphology. In what follows we will see which predictions are in accordance with the empirical facts.

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 differences 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 (2).[3] 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.

(2) 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:

[3] The other sentence type did not contain genitive-plurals, consider examples (6a) and (6b) from Lohmann & Conwell (2019): *(6a) The kids began a chat in front of the museum (of Natural History). (6b) The kids began to chat in front of the museum (of Natural History).*

> 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.

Recordings with hesitations, reading errors, false starts etc. were excluded. Four of the 82 participants had to be excluded altogether due to frequent disfluencies in their recordings. In the final data set, participants read 5.9 plurals and 5.3 genitive-plurals on average (between 3 and 7 plural forms and between 2 and 7 genitive-plurals). 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, with one lexeme, *actor*, featuring in two 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 token 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 using the 'U.S. English' setting (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 for durational differences between the relevant forms 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. Since we are interested in effects at the level of the morpheme as well as at the level of whole word, we fitted models with the dependent variable duration of S, as well as models with the duration of the whole word as the dependent variable.

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) on the whole corpus. 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 other 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. Phonetically voiced fricatives are shorter than unvoiced ones (e.g. Klatt 1976). Speakers often devoice final fricatives even when these fricatives are underlyingly voiced. We therefore decided to use a phonetic rather than a phonemic measurement for this variable, following the same procedure as implemented in Plag et al. (2017) and Tomaschek et al (2019): 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.
- SPEECHRATE. Segment durations become shorter with increasing speech rate. A frequently used measurement for speech rate is the number of segments (in the citation form) divided by the duration of a relevant linguistic unit. For our purposes, we computed 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.[4]

- 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 (with the values female and male) as a co-variate.

The regression models were fitted using the packages lme4 (Bates et al. 2014) and lmerTest (Kuznetsova et al. 2017) in 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). Variables were then eliminated following standard stepwise elimination procedures (e.g. Baayen 2008). A variable (fixed or random) was kept in the model if its presence (vs. its absence) led to a decrease in the AIC and to a significant improvement (p<0.05) in model fit tested via a log-likelihood test. If a variable did not meet these criteria, it was eliminated. To ensure a consistent elimination procedure we used the step function of the MASS package (Venables & Ripley 2002). 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.[5]

3.4 Collinearity

One issue we had to address is collinearity (e.g. Tomaschek et al. 2018). SPEECHRATE 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 expected since SPEECHRATE was computed with the number of segments in the base, and

[4] Due to the fact that the data come from a rhotic dialect of English, base-final /r/ was counted as a consonant.

[5] 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. With regard to the random effect structure we opted for a strategy that balances the risks of Type I errors and power (see Matuschek et al. 2017).

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 tested 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. We therefore included only SPEECHRATE in the intial models. 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 (but not NUMBEROFCONSONANTS) in the word duration models.

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*, $p<2.2e\text{-}16$, Wilcoxon test). Figure 1 shows the distribution by lexeme. Each pair of dots connected by a line represents one lexeme, with its two forms.

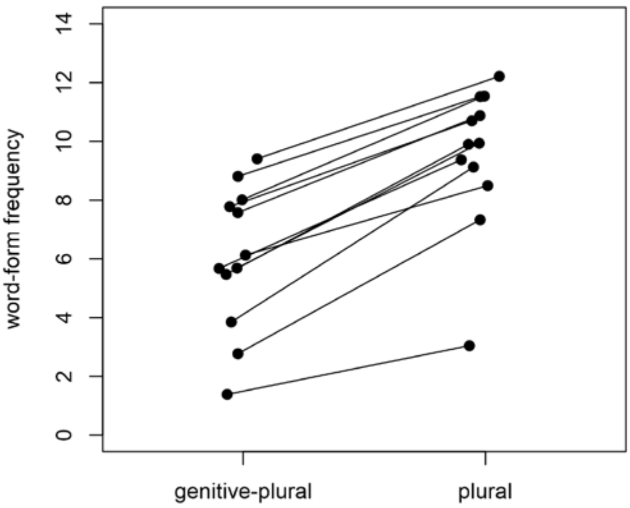

Figure 1: Word-form frequency by morphological category

The fact that LOGWORDFORMFREQUENCY and MORPH co-vary significantly (indicated by a correlation coefficient of *rho*= 0.66, $p<2.2e\text{-}16$, Spearman) precludes including them both in one model. We therefore fitted each model with only one of the two.

All models needed trimming of the residuals as the final stage of the model fitting process (see, for example, Baayen 2008: chapter 6 for discussion of model criticism in regression analysis). 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.

3.5. Transformation of the dependent variable and trimming of data sets

Before fitting regression models to the data we inspected the distribution of the durations of S and of the word durations. The non-normal distribution with several outliers suggested some trimming or transformation of these variables (see, for example, Baayen & Milín 2010 on issues of data trimming prior to analysis). We implemented different procedures, resulting in four slightly different data sets for final S, and four slightly different data sets for word duration. With each of the eight data sets we fitted two models, one model with MORPH as the variable of interest, the other model with LOGWORDFORMFREQUENCY as the variable of interest. This is resulted in 16 different models. The results of the models were highly similar to each other. For ease of exposition we therefore report only one final model for each combination of dependent variable (duration of S or word duration) and variable of interest (MORPH or LOGWORDFORMFREQUENCY). The models that we report are the ones in which the dependent variable is log-transformed and data points smaller or larger than 2.5 standard deviations are removed. The two data sets contain 860 observations for the duration of S, and 869 observations for word duration.

The generation of the eight data sets is documented in Appendix B, and the 16 final models are documented in Appendix C. The data sets 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. Results: The duration of plurals and genitive-plurals

4.1 Overview

Figure 2, left panel, 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 statistically significant (Wilcoxon test, *W*=110710, *p*=0.00013). The right panel shows the difference in duration between the plural word-forms and the genitive-plural word-forms. Genitive-plural word-forms are about 24 ms longer on average than the correponding plural word-forms (496 ms vs. 472 ms overall duration). This difference is statistically significant (Wilcoxon test, *W*= 107880, *p*=0.0021).

These may already be interesting results, but given the many potentially intervening influences described in the previous sections, these influences should be controlled for in a multivariate analysis, such as the mixed effects regression analysis described above.

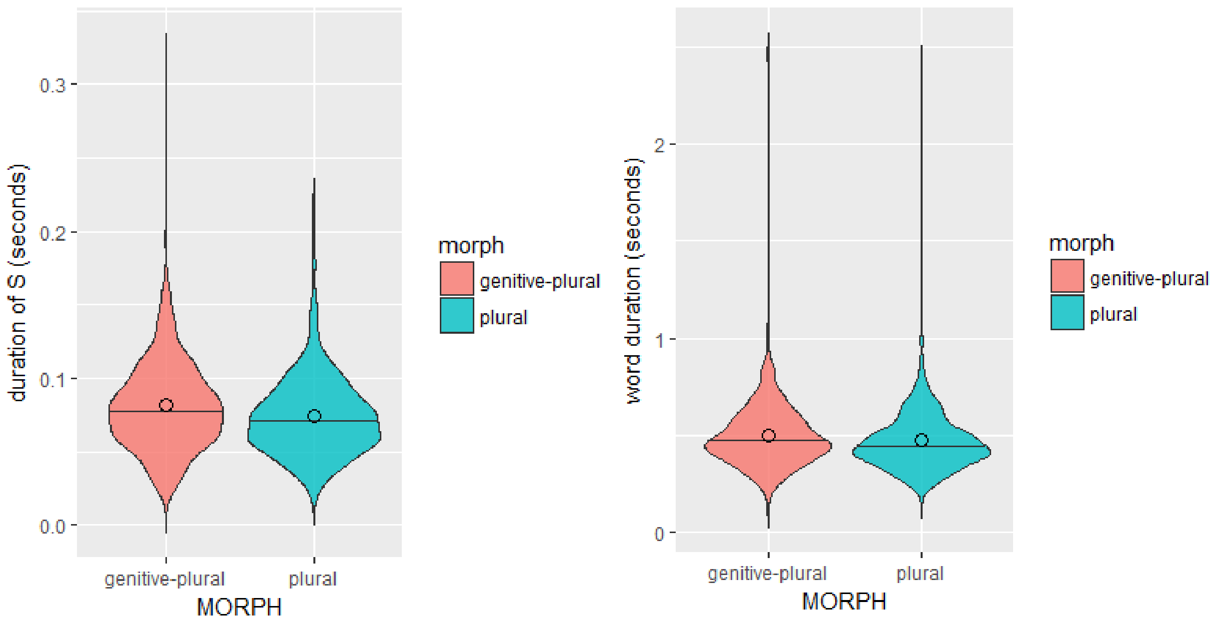

Figure 2: Durations of plural and genitive-plural S (left panel) and of plural and genitive-plural word-forms (right panel). The horizontal lines indicates the medians, the dots represent the mean.

4.2 The effect of morphological category: MORPH as variable of interest

Models were fitted according to the procedures described in section 3. The final model fitted to S durations contained three significant fixed effects, and two random effects. Similar results were obtained for the final model fitted to word durations, which contained four significant fixed effects, and three random effects. Table 2 gives an overview of the two models.

Table 2: Mixed effects regression models for the log-transformed duration of S and log-transformed word duration, with MORPH as variable of interest. (‘***’<0.001, ‘**’<0.01, ‘*’<0.05). For the fixed effects, the table gives the coefficients, standard errors are given in parentheses.

	S duration	**word duration**
Fixed Effects		
(Intercept)	-2.4387***	-0.7616***
	(0.0701)	(0.1528)
MORPHplural	-0.0936***	-0.0240***
	(0.0176)	(0.0053)
VOICINGvoiced	-0.1486***	-0.0218***
	(0.0229)	(0.0051)
SPEECHRATE	-0.0070*	-0.0566***
	(0.0034)	(0.0009)
NUMBEROFSYLLABLES		0.3637***
		(0.0677)
Random Effects		
Num. obs.	816	832
Num. groups: SUBJECT	78	78
Num. groups: ITEM	12	12
Var: SUBJECT (Intercept)	0.0065	0.0092
Cov: SUBJECT (Intercept) MORPHplural		-0.0010
Var: ITEM (Intercept)	0.0324	0.0407
Var: Residual	0.0606	0.0029

Let us first look at the fixed effects. The variable of interest, MORPH, is significant in both models: Genitive-plural forms are longer. The estimated mean durational difference between plural S and genitive-plural S is 7.8 ms, and the estimated mean durational difference between plural word-forms and genitive-plural word-forms is 14 ms. This means that both stem durations and suffix durations vary by morphological category.

Figure 3 shows the effect of MORPH on the duration of S (left panel) and on word duration (right panel). The lines show the predicted means, the dots represent the observed values.

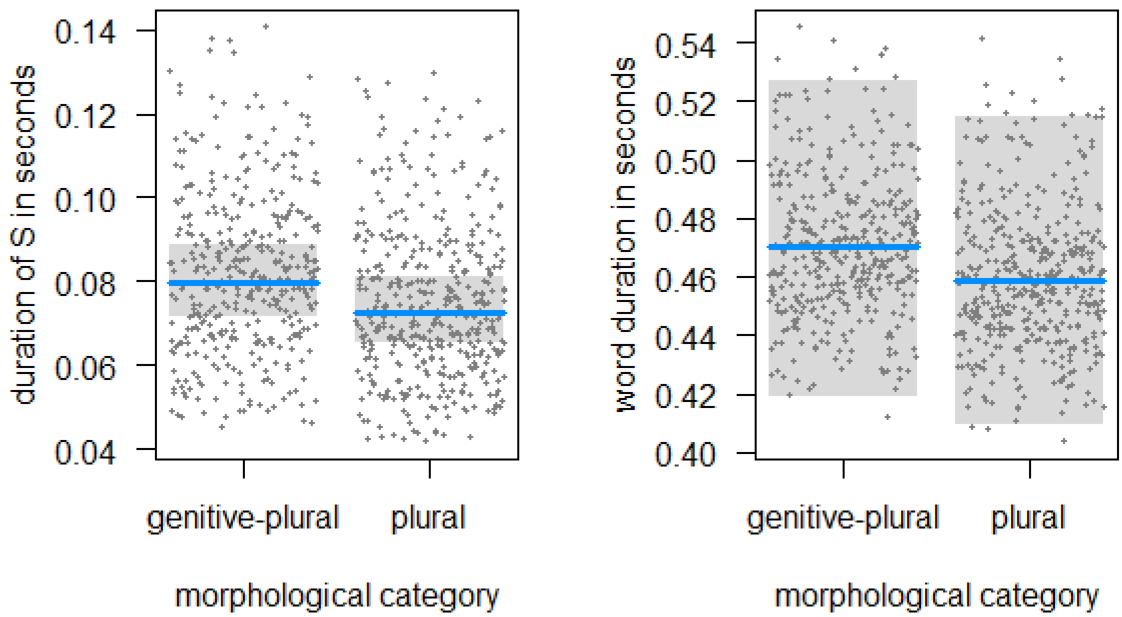

Figure 3: Back-transformed duration of S (left panel) and word duration (right panel) as predicted by the mixed effects regression models shown in Table 2.

In order to assess the relevance of the findings, it seems worthwhile to get a better idea of the strength of the effect. In particular, it seems worthwhile to compare the effect of the morphological category on the duration of S with that of voicing of S, which is a well-established phonetic parameter with an undisputed phonological correlate. This comparison will allow us to see whether the durational difference related to morphological category is perhaps negligible, even if statistically significant. To assess this, we standardized these fixed effects variables by subtracting the mean and dividing it by two standard deviations (see Gelman & Hill 2006: 56f for discussion), and then ran the model for the duration of S with these standardized predictors. The standardized coefficient of VOICING turned out to be stronger, but in the same range as the one for MORPH (0.106 for VOICING and 0.083 for MORPH).

This means that the effect of morphological category cannot be dismissed as being negligible.

In both models there are significant effects of VOICING and SPEECHRATE. The word duration model shows a significant also of NUMBEROFSYLLABLES. The effects of these control variables are as expected. The mixed effect structure is also of interest. The random intercept for item in both models shows that individual words vary in the durations of the S. Furthermore, individual speakers also show differences. With regard to the duration of S, speakers vary in how long they pronounce final S, which is not surprising. In the model for word duration a more complex random effect structure was justified, i.e. the inclusion of a random contrast for MORPH by SUBJECT. This means that there is some evidence that individual speakers vary in the way the durations of plurals and genitive-plurals differ: 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.

4.3 The effect of frequency: LOGWORDFORMFREQ as variable of interest

Models were fitted according to the procedures described in section 3. We fitted a model to the durations of S, and another model to the word durations. The results are similar. Table 3 gives an overview of the two models.

Table 3: Mixed effects regression models for the log-transformed duration of S and log-transformed word duration, with LOGWORDFORMFREQ as variable of interest. (‘***’<0.001, ‘**’<0.01, ‘*’<0.05). For the fixed effects, the table gives the coefficients, standard errors are given in parentheses.

	S duration	**word duration**
Fixed Effects		
(Intercept)	-2.2399***	-0.7050***
	(0.0883)	(0.1590)
VOICINGvoiced	-0.1332***	-0.0198***
	(0.0243)	(0.0052)
SPEECHRATE	-0.0126***	-0.0574***
	(0.0034)	(0.0010)
NUMBEROFSYLLABLES		0.3581***
		(0.0703)
LOGWORDFORMFREQ	-0.0230***	-0.0062***
	(0.0051)	(0.0011)
Random Effects		
Num. obs.	837	833
Num. groups: SUBJECT	78	78
Num. groups: ITEM	12	12
Var: SUBJECT (Intercept)	0.006	0.0003
Var: ITEM (Intercept)	0.051	0.0439
Var: Residual	0.070	0.0032

In both models the effect of LOGWORDFORMFREQ is statistically significant. With increasing word-form frequency the duration of S and the word duration become shorter. The effects of control variables and random effects are very similar to the ones we saw in the models with MORPH as variable of interest. Figure 4 illustrates the effect of LOGWORDFORMFREQ on the duration of S (left panel) and on word duration (right panel). The lines show the predicted means, the dots represent the residuals.

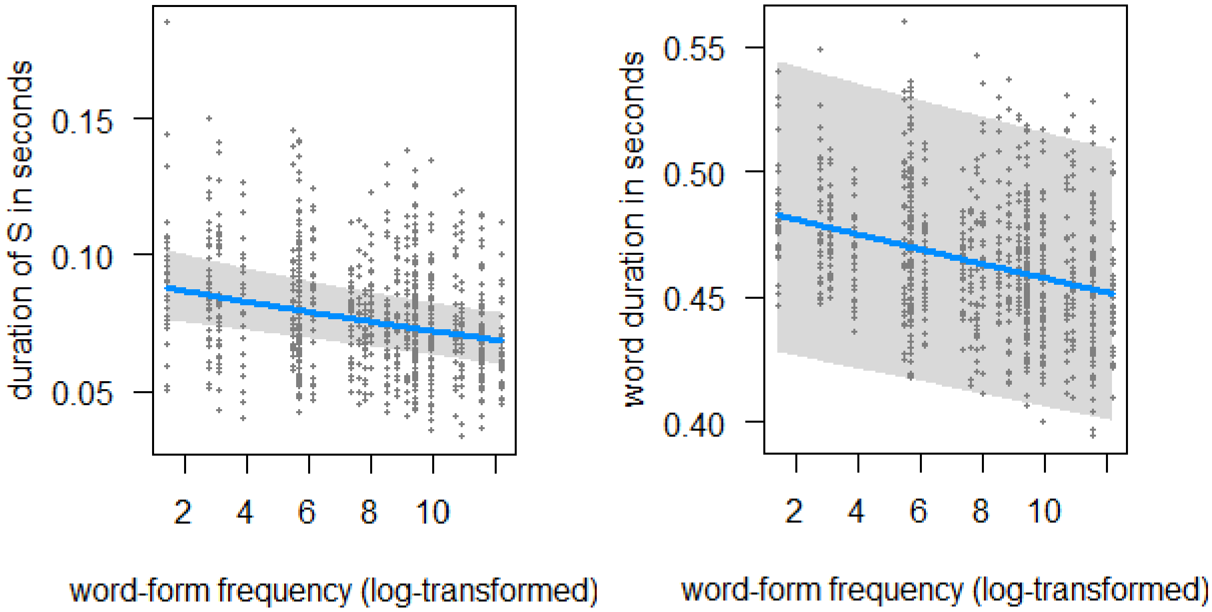

Figure 4: Back-transformed duration of S (left panel) and word duration (right panel) as predicted by the mixed effects regression models shown in Table 3.

5. Discussion

5.1. Predictions and outcomes

The statistical analysis with morphological category as the crucial predictor has shown that plural S and genitive-plural S differ significantly in duration, with the genitive-plural S being about 8 ms longer. However, the durational difference between plurals and genitive-plurals is not restricted to the final S. The duration of the whole word-form also varies by morphological category. Plural word-forms are 14 ms shorter than genitive-plural words, on average, which means that the effect of morphological category is spread over stem and suffix.

These results are not in accordance with the predictions of traditional structuralist-modular linguistic theories. Given previous research as summarized in section 2, the present study is not the first to find phonetic correlates of morphological structure that challenge these theories. The categories showing such effects are numerous, for instance the prefixes *mis-* and *dis-* in English (Smith et al. 2012), monomorphemic stems against suffixed stems (Sugahara and Turk 2004, 2009, Sproat 1993, Sproat and Fujimura 1993, Lee-Kim et al. 2013, Seyfarth et al. 2017, Plag et al. 2017, Tomaschek et al. 2019) and concern different phonological phenomena, such as gemination and degemination (Ben Hedia & Plag 2017, Ben Hedia 2019), /l/-velarization (e.g. Sproat 1993, Sproat and Fujimura 1993, Lee-Kim et al. 2013), as well as phonetic parameters such as vowel formants (Cohen 2015) or acoustic duration (e.g. Plag & Ben Hedia 2018). More recent versions of Lexical Phonology have tried to address such findings (e.g. Bermúdez-Otero 2015), but it is unclear whether, or how, the proposed amendments to this theory are able to accommodate the present findings.

In order to discuss our findings in the light of speech production models, it is useful to first take a look at the effect of word-form frequency that we found in our data. The analyses with word-form frequency as the crucial predictor demonstrated that higher word-form frequencies go together with shorter S durations and with shorter word durations. This is in line with the predictions of interactive theories. But how does this finding relate to the effects of morpho-syntactic category just discussed? 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 form *boys* has a log word-form frequency of 10.7, while the plural form *dogs* 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 very 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.

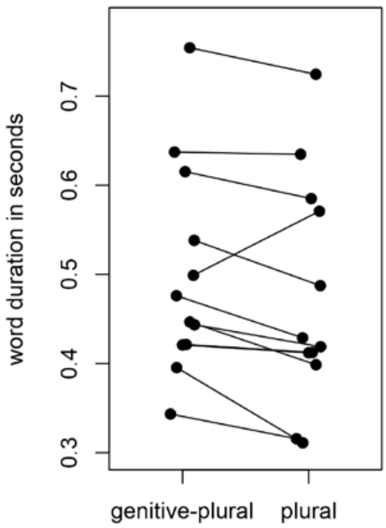

Figure 5: Word duration by morphological category

We see that for all pairs but one the genitive-plural form is of longer duration than the corresponding plural form (only the lexeme *patient* shows the opposite behavior, for unclear reasons). 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.[6]

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 result is expected by interactive theories (e.g. Goldrick et al. 2011) in which the strength of lexical activation may influence articulation. The word-form frequency effect found in the present study is also in accordance 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.

It should be noted, however, that the exact mechanisms by which higher frequency leads

[6] 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'*. For instance, the plural *dogs* has a frequency of 19889, and the genitive-plural *dogs'* has a frequency of 238. The plural *dogs* thus has a frequency ratio of 19889 / 238 = 83.6, while the genitive-plural has ratio of 238 / 19889 = 0.014. Taking the log of these frequency ratios as predictor or interest (instead of MORPH or LOGWORDFORMFREQ) yields a significant effect of the frequency ratio on the duration of S and the duration of the whole word in the expected direction. Larger frequency ratios go together with shorter durations. The models are included in Appendix C

to shorter durations is still to be worked out. Why speakers slow down when producing forms of lower frequency may have various origins and it is presently unclear which mechanisms contribute to it. In general it seems that items of lower frequency exhibit enhanced phonetic processing. Bell et al. (2009) suggest that longer durations for less frequent words result from lower lexical activation, which in turn leads to slower retrieval from the lexicon, which slows down articulation. With our data the situation is more complicated, however, as we are dealing with morphologically complex words. One might speculate that forms with more complex morpho-syntactic feature specifications also enhance processing costs, and therefore slow down articulation, which would result in the same effect. This makes it hard to tease apart the effect of word-form frequency from that of morphological complexity.

What seems clear is that strictly modular feed-forward models do not predict the patterning of our data. Furthermore, to account for word-form-specific frequency effects an architecture is necessary that allows for some kind of representation of inflected word-forms in the mental lexicon. More work is obviously needed in order to integrate the diverse findings in more comprehensive and more satisfactory models of speech production.

5.2 Morphological theory

Our results have implications for morphological theory. Word-form frequency effects for regularly inflected words in speech production are at odds with theories in which only morphologically irregular words, or highly frequent regular words, are assumed to be stored (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 can 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).

5.3 Alternative explanations and intervening factors

The results presented in this paper might also be resulting from some factors that we have not yet discussed. Two factors come to mind, prosody and spelling.

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 kinds of S 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 plural S. The opposite is the case in our data. Therefore, domain-final lengthening is unlikely to be the source of the durational differences between plurals and genitive-plurals that we observe.

Another potential influence is spelling. There are studies on the relationship of orthography and acoustic duration that have found that the number of orthographic symbols representing a sound correlates with the duration of that sound in speech (Brewer 2008). This would be in line with our results since the genitive-plural is represented by two orthographic symbols, <s> and <’>, the plural only by one, <s>. Other studies, however, failed to find this effect. For example, in Gahl’s study (2008) on heterographic homophones the covariate orthographic length did not reach significance, which means that homographs with longer spellings did not have longer durations in speech in the presence of other relevant variables (such as frequency).

There is also no theory available that may account for a possible correlation between number of orthographic symbols and acoustic duration, i.e. it is presently still unclear how orthographic effects on speech can be accounted for in a 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 the potential effects of these two factors.

5.4 Conclusion

To conclude, this article has shown that, phonetically, plurals and genitive plurals in English are not homophonous. Plurals are shorter than genitive-plurals, and this holds for stems and for the final S. The fact that complex words vary in their durational characteristics depending on their morphological make-up has implications for our thinking about lexical organization and lexical processing. 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

We thank Ariel Cohen-Goldberg for his extremely helpful comments on a previous version of this paper. 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.

Bermúdez-Otero, R. (2017). Stratal phonology. In S. J. Hannahs & Anna Bosch (eds.), *Routledge handbook of phonological theory*, 100–134. London, UK: Routledge.

Bermúdez-Otero, R.. (2015). Amphichronic Explanation and the Life Cycle of Phonological Processes. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 374–399. Oxford: Oxford University Press.

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.

Cohen, C. (2015). Context and paradigms: Two patterns of probabilistic pronunciation variation in Russian agreement suffixes. *The Mental Lexicon* 10(3). 313–338.

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

Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, 93, 283–321.

Engemann, U. M., I. Plag & J. Zimmermann. (2019). Morphological boundaries and stem duration in English: Replicating experimental results with corpus data. Poster presented at 12th Mediterranean Morphology Meeting, 27-30 June 2019, Ljubljana.

Fry, E. (2004). Phonics: A large phoneme-grapheme frequency count revised. *Journal of Literacy Research* 36(1). 85–98.

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

Goldrick, M., H. R. Baker, A. Murphy & M. Baese-Berk (2011). Interaction and representational integration: Evidence from speech errors. *Cognition* 121(1). 58–72.

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. *Journal of Phonetics*.

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.

Matuschek, H., R. Kliegl, S. Vasishth, H. Baayen & D. Bates. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315.

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. & S. Ben Hedia. (2018). The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration. In S. Arndt-Lappe, A. Braun, C. Moulin & E. Winter-Froemel (eds.), *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and Ludicity*. 93-116. Berlin, Boston: De Gruyter.

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

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.

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 (*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 (*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).

Appendix C

In the following tables, 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.

Table C.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***	-2.2399***	0.7268***	0.0836***	-2.4746***	-2.4275***	0.7098***
	(0.0050)	(0.0723)	(0.0701)	(0.0070)	(0.0049)	(0.0885)	(0.0883)	(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***	-0.1332***	-0.0135***	-0.0083***	-0.1454***	-0.1318***	-0.0129***
	(0.0017)	(0.0239)	(0.0229)	(0.0023)	(0.0020)	(0.0238)	(0.0243)	(0.0023)	(0.0016)	(0.0237)	(0.0243)	(0.0025)
SPEECHRATE	-0.0005*	-0.0081*	-0.0070*	-0.0010**	-0.0008**	-0.0085*	-0.0126***	-0.0011***	-0.0006*	-0.0081*	-0.0120***	-0.0012***
	(0.0002)	(0.0035)	(0.0034)	(0.0003)	(0.0003)	(0.0035)	(0.0034)	(0.0003)	(0.0002)	(0.0035)	(0.0034)	(0.0004)
LOGWORDFORMFREQ						-0.0235***	-0.0230***	-0.0022***				
						(0.0050)	(0.0051)	(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	287.2161	-3597.6545	-4090.2107	229.6394	289.8290	-3551.6202
BIC	-4045.4870	274.8226	197.2098	-3562.0178	-3929.9005	266.2152	320.3249	-3564.5960	-4038.4352	262.6809	322.9462	-3518.3706

Log Likelihood	2052.9244	-113.8821	-75.1394	1811.2607	1998.7754	-109.5910	-136.6081	1805.8272	2056.1053	-107.8197	-137.9145	1782.8101
Num. obs.	818	831	816	831	867	828	837	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	0.006	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	0.051	0.0005	0.0002	0.0335	0.0333	0.0003
Var: Residual	0.0003	0.0668	0.0606	0.0006	0.0005	0.0657	0.070	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 C.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