Homophony and morphology: The acoustics of word-final S in English[1]

INGO PLAG, JULIA HOMANN and GERO KUNTER

Heinrich-Heine-Universität Düsseldorf

(Received 5 September 2014; revised 5 May 2015)

Recent research has shown that homophonous lexemes show systematic phonetic differences (e.g. Gahl 2008, Drager 2011), with important consequences for models of speech production such as Levelt et al. (1999). These findings also pose the question of whether similar differences hold for allegedly homophonous affixes (instead of free lexemes). Earlier experimental research found some evidence that morphemic and non-morphemic sounds may differ acoustically (Walsh & Parker 1983, Losiewicz 1992). This paper investigates this question by analyzing the phonetic realization of non-morphemic /s/ and /z/, and of six different English /s/ and /z/ morphemes (plural, genitive, genitive-plural and 3rd person singular, as well as cliticized forms of *has* and *is*). The analysis is based on more than 600 tokens extracted from conversational speech (Buckeye Corpus, Pitt et al. 2007). Two important results emerge. First, there are significant differences in acoustic duration between some morphemic /s/'s and /z/'s and non-morphemic /s/ and /z/, respectively. Second, there are significant differences in duration between some of the morphemes. These findings challenge standard assumptions in morphological theory, lexical phonology and models of speech production.

1. INTRODUCTION

Recent studies on lexeme homophony have shown that seemingly homophonous words actually differ in phonetic details such as length and vowel quality (e.g. Gahl 2008, Drager 2011). Likewise, realizational differences can be found with constituents of complex words. Kemps et al. (2005a, b) showed that free and bound variants of a stem differ acoustically and that listeners make use of such phonetic cues in speech perception. Findings like these pose a challenge to traditional models of speech production which locate frequency information at

---

the level of the phonological form, and which postulate that phonetic processing does not have access to morphological information (e.g. Levelt & Wheeldon 1994, Levelt et al. 1999).

Homophony below the level of the lexeme, however, has not received much attention so far. There are two questions that warrant closer inspection. The first concerns seemingly homophonous segments that do, or do not, represent morphemes. For example, the morphemic /s/ in *laps* might acoustically differ from the non-morphemic /s/ in *lapse*, as suggested by Walsh & Parker (1983). The second question of interest concerns morphemes that share one phonological form while representing different meanings or functions. There could be fine phonetic differences between such morphemes, as for example between plural /s/ (as in *two heaps*) and 3rd person singular /s/ (as in *she keeps*).

In this paper, we investigate both questions using data from the Buckeye Corpus (Pitt et al. 2007), we look into the issue of morpheme homophony by investigating the phonetic realization of different English /s/ and /z/ morphemes, which can denote plural, genitive, genitive-plural and 3rd person singular, as well as cliticized forms of *has* and *is*. In addition, we compare non-morphemic final /s/ and /z/ with their morphemic counterparts. We show that there are significant differences between many of the different kinds of /s/ and /z/. Our regression models with a number of pertinent covariates (e.g. speech rate, position in the phonological phrase, frequency, etc.) demonstrate, for example, that plural and non-morphemic /s/ and /z/ are significantly longer than most other /s/ and /z/ morphemes. These findings are unexpected from a theoretical perspective and pose further challenges to extant theories of morphology and widely accepted models of speech production.

The paper is structured as follows. In the next section we lay out in more detail the general problem of the relationship between morphological structure and the phonetic signal, and we will develop our research questions. Section 3 introduces the methodology of our corpus-based study. The results are presented in Section 4, followed by a discussion and conclusion.

2. MORPHOLOGICAL STRUCTURE AND THE PHONETIC SIGNAL

Traditionally, morphemes are defined as linguistic units that consist of a meaning and a phonological form. In the case of allomorphy, one meaning is linked to several phonological forms whose choice depends on constraints that can be phonological, morphological or lexical in nature. The allomorphs /s/, /z/ and /ɪz/ of the English plural morpheme are a textbook example of such allomorphy. If the base noun ends in a sibilant, /ɪz/ is chosen, if the base ends in a non-sibilant voiceless consonant, /s/ is chosen, and /z/ occurs in all other contexts. This distribution is exactly the same for the English 3rd person singular marker, i.e. the two morphemes are identical at the form level, both in terms of their exponents and in terms of their distribution. /s/, /z/ and /ɪz/ are also allomorphs of the genitive marker (though with slightly different distributional constraints,

see Bauer et al. 2013: 69, 129, 145). The plural morpheme, the genitive-plural morpheme, the genitive morpheme and the 3rd person singular morpheme thus share homophonous exponents. Similarly, the cliticized forms of *has* and *is* can be either /s/ or /z/ depending on the presence of voicing in the context, which makes them homophones to each other and homophonous to the /s/ and /z/ allomorphs of the plural, 3rd singular and genitive morphemes (Bauer et al. 2013: 85). Crucially, there is nothing in the representation of the allomorphs that could cause systematic differences in phonetic implementation between the different morphemes, or between the same segments when they do not represent morphemes.

The said morphemes are treated in a similar way in standard feed-forward formal theories of morphology–phonology interaction (e.g. Chomsky & Halle 1968, Kiparsky 1982). In such models the allomorphy is determined at a particular phonological cycle inside the lexicon, and at the level of underlying representations. Once the right underlying form is derived, the morphological boundary of the respective cycle is erased (a process called 'bracket erasure', see Chomsky & Halle 1968, Kiparsky 1982) and the form leaves the lexicon. All further phonological processes are relegated to another module called 'post-lexical phonology' and later to the articulatory component, neither of which have access to morphological information. As in traditional structuralist accounts, there is nothing in the system that would allow for systematic phonetic differences between homophonous suffixes, or between morphemic and non-morphemic sounds.

In the framework of Prosodic Morphology, there is a complex mapping of morphological structure onto prosodic structure (e.g. Nespor & Vogel 2007), with the phonological or prosodic word as the central notion. The different types of mapping are responsible for various morpho-phonological alternations (such as assimilation or resyllabification; see Plag (2003) for an introductory treatment and exemplification). Since prosodic boundaries may have phonetic correlates, it has been shown in many studies that morphemes at such boundaries may show systematic differences in phonetic implementation that correlate with differences in their prosodic position (see, for example, Keating 2006). In this framework, any phonetic difference found between two phonologically homophonous affixes would therefore be derivable from a difference in the prosodic structure that goes with the two different affixes.

The distinction between lexical and post-lexical processes also features prominently in psycholinguistics. According to widely accepted models of speech production (e.g. Levelt et al. 1999), the aforementioned morphemes and clitics would not differ in their realization from corresponding non-morphemic /s/, /z/ and /ɪz/. In these models, meanings are stored in the mental lexicon, and their corresponding forms are represented phonologically. Thus, what is used as a basis for articulation is the phonological form only, and the module called 'articulator' does not have access to any information regarding the lexical origin of a sound. Leaving stylistic and accentual differences aside, a certain string of phonemes in a given context will therefore always be articulated in the same way, irrespective

183

of its morphemic status, and only modulo the phonetic variation originating from purely phonetic sources such as speech rate or context. It is yet unclear whether more nuanced versions of this model (such as Keating & Shattuck-Hufnagel 2002) could accommodate morpho-phonetic effects.

Recent research on the homophony of lexemes suggests that such a model of speech production may be insufficient. Gahl (2008) investigated the acoustic realization of 223 supposedly homophonous word pairs such as *time* and *thyme*, and found that, quite consistently, the more frequent member of the pair, e.g. *time*, is significantly shorter than the respective less frequent one, e.g. *thyme*. This can be taken as evidence that two homophonous lexemes cannot be represented exclusively by one identical phonological form with information on their combined frequency, but that the individual frequencies must be stored with the respective lemmas and have an effect on their articulation.

Similarly, Drager (2011) found that the different functions of *like* go together with different acoustic properties. Whether *like* is used as an adverbial, as a verb, as a discourse particle or as part of the quotative *be like* has an effect on several phonetic parameters, including the ratio of the duration of /l/ to vowel duration, the pitch level and the degree of monophthongization of the vowel /aɪ/. These fine differences indicate that at the phonetic level two or more phonologically homophonous lemmas may differ. These effects seem to hold also for function words, as shown in Lavoie (2002), who investigated *four* and *for* – Jurafsky et al. (2002) also found acoustic differences between the two words.

Below the word level, there is evidence that phonemically identical strings may systematically vary in their phonetic realization, depending on morphemic status. This seems to run counter not only to standard models of speech production but also to the structuralist and formal theories of phonology–morphology interaction. Kemps et al. (2005b) found that phonologically segmentally identical free and bound variants of a base (e.g. *help* without a suffix as against *help* in *helper*) differ acoustically. Furthermore, the authors showed that Dutch and German listeners do make use of such phonetic cues in speech perception (see also Kemps et al. 2005a).

In their experiments, Sugahara & Turk (2004, 2009) also found phonetic differences between the final segments of a monomorphemic stem as against the final segments of the same stem if followed by a suffix. Stems followed by certain suffixes had slightly longer rhymes than their monomorphemic counterparts.

There is also articulatory evidence on the variability of intergestural timing in monomorphemic and complex words which points at incongruities in the representations of homophones. In an electropalatographic study, Cho (2001) found that in Korean, timing of the gestures for [ti] and [ni] shows more variation when the sequence is heteromorphemic than when it is tautomorphemic, which indicates that morphological structure is reflected in the details of the articulatory gestures, with potential acoustic correlates in the speech signal.

The interpretation of these findings at the theoretical level is highly controversial, however. While some researchers (e.g. Kemps and colleagues, Jurafsky and

colleagues) would argue for the incorporation of phonetic detail into lexical, i.e. morphological, representations, other researchers try to explain the findings as reflexes of other, e.g. prosodic or contextual, properties (e.g. Sugahara & Turk).

In view of the controversial findings and implications, homophony below the level of the lexeme seems worthy of closer inspection. If there are fine differences between supposedly homophonous free lexemes and between bound and free realizations of the same stem, there could well be systematic differences between morphemic and non-morphemic sounds, and differences between supposedly homophonous affixes. Not much research has been done in this area, but some experimental studies are available that have looked at morphemic (i.e. affixal) versus non-morphemic (strings of) sounds.

Walsh & Parker (1983) carried out a production experiment with three homophonous word pairs and measured the length of /s/ in monomorphemic words and in words that were homophonous to the monomorphemic ones but contained a final morphemic /s/ (e.g. *lapse* versus *laps*). The experiment had three different conditions, and in each condition the word pairs were presented in a different context. The authors then compared the means of morphemic and non-morphemic /s/ across the three different conditions of their experiment. In two of the conditions there is a small difference of nine milliseconds in the means of the two different kinds of /s/, while the third condition shows no difference between morphemic and non-morphemic /s/. Based on these means the authors concluded that 'the durational differences of final /s/s observed in Conditions I and II appear to be a function of the morphological status of the /s/' (Walsh & Parker 1983: 204). Given the very small data set and other methodological problems such as the lack of any inferential statistical analysis, or the integration of any phonetic covariates, Walsh & Parker's results may be met with great scepticism.[2]

In a similar study, Losiewicz (1992) investigated the acoustic difference between morphemic, i.e. past tense, /d/ and /t/, and non-morphemic /d/ and /t/, and also found durational differences between the two sets of sounds. As Hanique & Ernestus (2012) point out, however, Losiewicz's study suffers from serious methodological problems, such as very small data sets and the use of insufficient frequency measures. Another problem that is not mentioned by Hanique & Ernestus is that Losiewicz tested both /t/ and /d/ without including voicing as a covariate.

Li et al. (1999) investigated child-directed speech and found that plural -*s* was longer than 3rd person singular -*s*, but they attributed this difference to the fact that plural nouns had a greater tendency to appear in utterance-final position in their data set, such that the effect could be conceived as by-product of utterance-final lengthening. In a more recent experimental study of mothers' and young children's speech, Yung Song et al. (2013) looked at the duration of /z/

---

[2] Another study of morphemic versus non-morphemic /s/ is Schwarzlose & Bradlow (2001), but this is only available as a one-page abstract and is therefore of little use.

185

in four monosyllabic plural forms as against two monosyllabic 3rd person singular forms and three non-morphemic forms. They found that in the mothers' speech morphemic /z/ was longer than non-morphemic /z/, but the effect was restricted to utterance-final position. There was no significant difference between plural and 3rd person /z/.

Baker et al. (2007), also using experimental data, found acoustic differences (in durational and amplitude measurements) between morphemic and non-morphemic initial *mis-* and *dis-* (as in, e.g., *distasteful* versus *distinctive*). Again there are methodological problems, such as the fact that morphemic and non-morphemic strings did not only differ in morphemic status but also in phonological properties that may have directly affected their acoustic realization, such as stress. For example, *distasteful* may have a secondary stress on the first syllable (perhaps with additional differences in vowel quality) while *distinctive* may not, which may have an effect on duration and amplitude.

In sum, there is some evidence that there might be systematic duration differences between affixes and the corresponding homophonous non-morphemic sounds, but the data sets are very small and do not represent natural conversational speech, and the effects found are not always convincing due to methodological shortcomings. Nevertheless the previous results are promising enough to warrant further inquiry into the homophony of morphemic and non-morphemic sounds.

With regard to the acoustic differences between different homophonous suffixes, we have to say that to the best of our knowledge, this question has never been systematically investigated, although it potentially has important theoretical implications, as any effect found in this area would seem to run counter to established theories of morphology–phonology interaction and models of speech production. Systematic morpho-phonetic effects would raise the question of the place of phonetic detail in lexical representation and lexical processing.

A study is therefore called for that investigates the two aspects of suffix homophony on a larger scale, preferably using data from natural conversations. This paper presents such a study, testing the two null hypotheses given in (1) and (2).

(1)   Null Hypothesis 1
       There is no durational difference between non-morphemic and morphemic sounds (e.g. between plural /s/ and non-morphemic word-final /s/).

(2)   Null Hypothesis 2
       There is no durational difference between homophonous suffixes (e.g. between plural /s/ and 3rd person singular /s/).

3. METHODOLOGY

3.1 *Data*

Let us first look at morphemic /s/ and /z/. We investigate six morphemes that share the allomorphs /s/ and /z/, namely the 3rd person singular marker, the plural marker, the genitive marker, the combined genitive-plural marker, the cliticized form of *has* and the cliticized form of *is*. We focus in this paper on the duration of the two allomorphs /s/ and /z/. The allomorph /ɪz/, which is restricted to plural, genitive and 3rd person singular marking, is not considered.[3]

A note on our terminology is in order. We use capitalized 'S' as an umbrella term for the two segments /s/ and /z/ in word-final position. The term 'morphemic S' is used as an umbrella term for the clitics and suffixes. Furthermore, we use the term 'base' both for morphological bases as well as for hosts in cliticization, and for the string of sounds that precedes the final S in monomorphemic words.

The data source for this study is the Buckeye Corpus of Conversational Speech (Pitt et al. 2007). This corpus comprises about 300,000 words from 40 long-time local residents of Columbus, Ohio, who were recorded conversing freely with an interviewer for about one hour each. In addition to the raw speech files, the Buckeye Corpus offers time-aligned written and phonetic transcriptions of the interviews.

Why do we use conversational speech instead of experimental data? While experimental data may provide a better opportunity to control potentially intervening variables in various ways, data obtained in this way sometimes also raise concerns about their validity. It has been shown, however, that, primarily due to modern statistical techniques such as mixed-effects regression (Baayen et al. 2008), more natural data can be fruitfully employed to investigate issues of morpho-phonetic detail by including pertinent covariates that control for many sources of variability in the data (see, for example, Ernestus & Warner (2011) for an overview). We therefore opted for conversational data.

Examples of the different kinds of morphemic S from our data set are given in (3).

(3)   (a)   It *feels* like they're trying to tease me (*3rd person singular*)

      (b)   A grand total of six *years* as an undergrad (*plural*)

      (c)   Write a short story in that *author's* style (*genitive*)

      (d)   The local senior *citizens'* home (*genitive-plural*)

      (e)   That's the way my *family's* always been (has *clitic*)

      (f)   My *brother's* twenty four (is *clitic*)

---

[3] Another potential candidate is cliticized *us*, as in *let's*. As this form does not show any variation with regard to its host (it is always *let*), a comparison with the other /s/ and /z/ morphemes in a mixed-effects regression model is problematic. Cliticized *us* was therefore excluded from this study.

For the analysis of length differences between different types of morphemic S we needed items that consisted of a base and of one of the types of morphemic S introduced above. While the clitics can take all sorts of bases, affixal S is more restricted in this respect. To keep the set of items as homogeneous as possible, only items with verbs, nouns and pronouns (indefinite and personal) as bases entered the data set.

Using the POS-tagged orthographic transcription of the Buckeye corpus, 100 tokens of each morphemic S were randomly extracted. If a type occurred more than 12 times, all additional tokens were replaced by other, randomly sampled types. If less than 100 tokens were available for a certain kind of morphemic S, all available tokens were extracted. We limited the amount of data to these numbers due to the very time-consuming manual inspection and readjustment of the automatic segmentations of the Buckeye Corpus (see Section 3.2.1 for details). The overall set of morphemic items amounted to 460 (i.e. tokens), representing 293 types. Of these items, 11 were excluded because acoustic and visual inspection either revealed that the morphemic S was not realized as [s] or [z] in the speech signal (but, e.g., as [ɪz] or [ʃ], or omitted completely), or the final S was not unambiguously attributable to the item due to assimilation to an initial sibilant in the following word. There was also one case in which the speaker purposefully lengthened the S for stylistic reasons. After inspection of the distribution of the duration measurements we also excluded as outliers three items that were longer than 250 ms. Eventually, 448 items entered the acoustic analysis of morphemic S.

In order to investigate the potential difference between morphemic and non-morphemic S, we also sampled a set of non-morphemic word-final S tokens from the corpus. This sample was created as follows. In an initial step we extracted all words from the corpus that ended in [s] or [z], irrespective of their expected standard pronunciation. Multimorphemic words, irregular 3rd person singular forms (i.e. *does*, *has* and *is*), pronouns and determiners were manually removed from this initial set. If a monomorphemic word, i.e. type, was produced more than once by a speaker, only one randomly selected token of that type entered the pre-final data set ($N = 3057$ tokens).[4] To arrive at a reasonably sized set of data, 240 words were randomly sampled from the pre-final data set. Thirty-four items with anomalies in the signal (similar to those mentioned in the previous paragraph for morphemic S) were excluded, as were seven outliers that were longer than 250 ms. The final data set consisted of 199 words with non-morphemic S, about half the number of items that was in the set of morphemic S.

In the overall final data set, the different types of S were distributed as shown in Figure 1.

---

[4] We use the notion of 'type' here to refer to a specific pronunciation of a word as given in the Buckeye transcriptions. That is, if a given word was pronounced in different ways, for example with variable deletion of a particular segment, these were treated as different types.

188
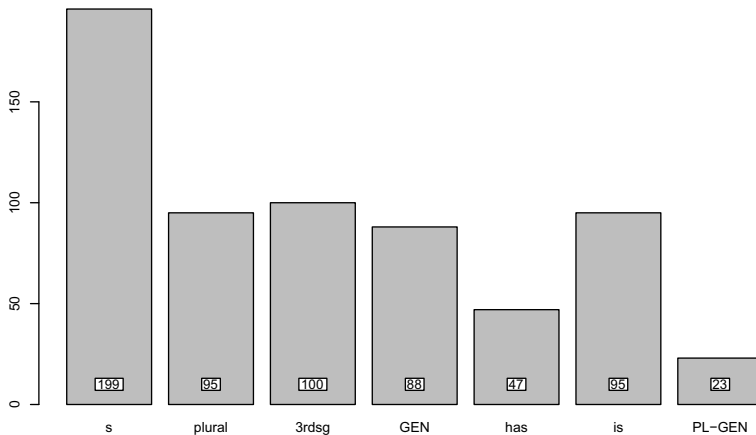
Figure 1
Distribution of different types of S in the data set. (Abbreviations: s = non-morphemic S,
3rdsg = 3rd person singular, GEN = genitive, PL-GEN = genitive-plural).

3.2 *Analysis*

3.2.1 *Acoustic measurements*

With the help of LaBB-CAT (Fromont & Hay 2008), a freely available speech
corpus management system formerly known as ONZE Miner, the Buckeye
transcripts were converted into textgrid files. These files could then be used for
further segmentation and analysis with the help of the acoustic analysis software
Praat (Boersma & Weenink 2013).

Buckeye's (partly automatic) phonetic annotations were manually checked for
each item and adjusted where necessary. Boundaries marking the beginning of an
item or of an S were moved to the zero crossing that was closest to the point
where both spectrogram and waveform indicated the initiation of the gesture
for the respective segment, i.e. in the case of S, the boundaries were set to
the zero crossing closest to the onset of the friction visible in the waveform.
Boundaries marking the end of an item and thus the end of an S were moved
to the zero crossing closest to the point where the initiation of the gesture for the
following segment became visible in both spectrogram and waveform. In cases
with no following segment, the boundary was set to the point where the friction
of the S dropped to silence. After manual checking and adjustment of all relevant
intervals, relevant acoustic measurements such as length and voicing (for details,
see Section 3.2.2) were taken automatically with the help of a Praat script.

We wanted to model the length of S both in absolute terms and in relative terms,
i.e. in relation to the length of its base. Studies of geminates (e.g. Oh & Redford
2012) have shown that differences in phonetic duration can be meaningfully
interpreted as relative or absolute, depending, among other things, on the language
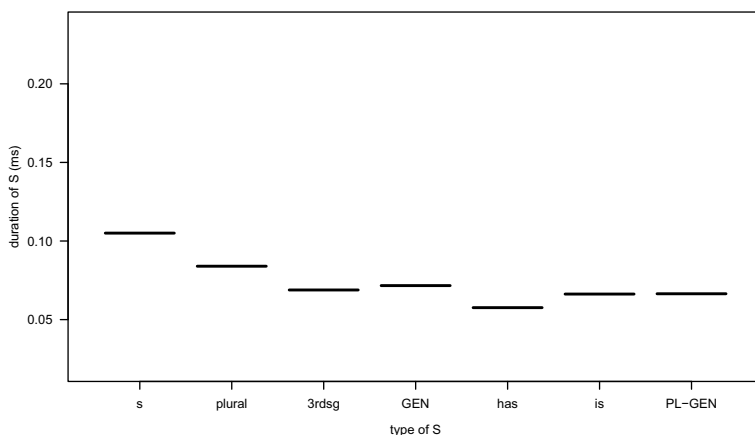
Figure 2
Duration of different types of S in the data set. (Abbreviations: s = non-morphemic S,
3rdsg = 3rd person singular, GEN = genitive, PL-GEN = genitive-plural).

under investigation. Given that very little is known about the absolute or relative length of affixes, it seemed reasonable to test both kinds of dependent variables. Therefore, the absolute duration of S in a given token was obtained from the duration of the segment in milliseconds, and the relative duration was calculated by dividing the absolute duration of the S by the duration of the whole word, i.e. by the sum of base duration and S duration.

The distribution and means of the duration measurements by type of S are given in Figure 2. Each dot represents one measurement and the lines indicate the means.

The distribution of the measurements shows rather clear differences between the different types of S, and an ANOVA shows a significant effect of TYPE OF S (F = 20.196, p < 2.2e–16). The results of pair-wise comparisons of the means using Tukey contrasts are summarized in Table 1.[5] We find ten significant contrasts. Non-morphemic S differs from all morphemic S's, and there are four significant differences among the different types of morphemic S.

While this may already look like a very interesting result, it has to be treated with great caution. Previous studies of the acoustic duration of words and individual sounds have shown that this acoustic parameter is subject to various acoustic and non-acoustic factors, such as speech rate, frequency, final lengthening or the phonological context of the surrounding segments. Variable durations for the same word type may also arise from more general processes of

---

[5] For the comparisons the packages `multcomp` (Hothorn et al. 2008) and `sandwich` (Zeileis 2004, 2006) were used, implementing the specific procedure suggested by Herberich et al. (2010) for scenarios with unbalanced group sizes, non-normality and heteroscedasticity.

190

	Estimate	Std. Error	*t*-value	Pr (> |*t*|)	
plural - s	−0.0210570	0.0052977	−3.975	0.00154	**
3rdsg - s	−0.0361719	0.0046476	−7.783	<0.001	***
GEN - s	−0.0333925	0.0045509	−7.337	<0.001	***
has - s	−0.0474129	0.0042647	−11.118	<0.001	***
is - s	−0.0387739	0.0038523	−10.065	<0.001	***
PL-GEN - s	−0.0385812	0.0049358	−7.817	<0.001	***
3rdsg - plural	−0.0151149	0.0056085	−2.695	0.09851	.
GEN - plural	−0.0123355	0.0055286	−2.231	0.27464	
has - plural	−0.0263559	0.0052955	−4.977	<0.001	***
is - plural	−0.0177169	0.0049694	−3.565	0.00697	**
PL-GEN - plural	−0.0175242	0.0058495	−2.996	0.04356	*
GEN - 3rdsg	0.0027794	0.0049092	0.566	0.99761	
has - 3rdsg	−0.0112410	0.0046450	−2.420	0.18775	
is - 3rdsg	−0.0026020	0.0042695	−0.609	0.99640	
PL-GEN - 3rdsg	−0.0024093	0.0052680	−0.457	0.99929	
has - GEN	−0.0140204	0.0045483	−3.083	0.03370	*
is - GEN	−0.0053814	0.0041641	−1.292	0.85201	
PL-GEN - GEN	−0.0051887	0.0051829	−1.001	0.95232	
is - has	0.0086390	0.0038492	2.244	0.26811	
PL-GEN - has	0.0088317	0.0049334	1.790	0.54849	
PL-GEN - is	0.0001927	0.0045816	0.042	1.00000	

Table 1
Multiple comparison of means of duration of S (Tukey contrasts). (Significance codes:
'***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.)

phonetic reduction, which may affect duration but also non-durational parameters such as vowel quality or number of segments and syllables. In any study interested in one particular factor, in our case the influence of morphological status, these other influences need to be controlled, for example as covariates in a regression analysis.

3.2.2 *Covariates*

The specific set of covariates chosen for the present study is very similar to that of other studies that have investigated duration effects in morphologically complex words using speech corpus data (for example Pluymaekers et al. 2005b, 2010; Hanique et al. 2013). In the following we will briefly discuss each of the covariates, starting with phonetic covariates, followed by covariates encoding lexical properties, and finally context-encoding variables.

191

Local speech rate. An obvious variable that needs to be controlled for is speech rate. We used two measures of speech rate. As one measure we calculated the speech rate in the discourse around each item (i.e. the 'local' speech rate in syllables per second). This was done by counting the number of syllables in the context of up to ten seconds immediately preceding and succeeding the item, if not interrupted beforehand, e.g. by a pause or by the interviewer. The number of syllables was then divided by the number of seconds of context that was considered. The inspection of the distribution of this variable showed two clear outliers with speech rates greater than 15 syllables per second. Furthermore, there was one item for which there were no surrounding syllables, so that no speech rate could be computed. Those three items were removed from the data set before statistical modeling, leading to a reduction of the set of items with non-morphemic S from 199 to 196.

Base duration. With regard to an even more local speech rate, we measured the length of the base. Other things being equal, a longer duration of the base will indicate a slower speech tempo. This measurement is also able to at least partially control for the lengthening effect of accentuation: as Turk & White (1999) show, word-final (non-morphemic) consonants are significantly longer if the word they belong to is accented.

Voicing. We included voicing to account for the effect of allomorphy. It is also well known that voicing affects the length of fricatives, with voiced fricatives being shorter (e.g. Klatt 1976). Based on an inspection of the distribution of the measurements taken by the algorithm used in Praat, an S was considered to be voiced if the algorithm could detect a periodic pitch pulse in more than 75 percent of the overall duration of the segment.

Number of syllables. We also included another measure of length, the number of syllables. As shown in Lindblom (1963) for Swedish vowels, and Nooteboom (1972) for Dutch vowels, vowel segments may tend to be shorter if they are followed by more syllables (but see Hanique et al. (2013) for somewhat different findings). It is unclear whether this effect can also be found in English, and whether it pertains to fricatives, but it seemed safe to include this covariate nevertheless. In any case, this effect can be conceptualized as a kind of compression effect, where words with more syllables undergo reduction. We used two kinds of syllable count. The first one was to extract syllable counts from CELEX (Baayen et al. 1995) where possible; in all other cases the number of vowels (or diphthongs) in a word was taken as an indication of its number of syllables. The second one was to count the syllabic nuclei that were actually pronounced.

Number of consonants immediately preceding S. Another factor influencing the length of consonantal segments is their occurrence in a cluster. Consonants in clusters tend to be shorter (e.g. Klatt 1976). We therefore included the number of consonants preceding S as a covariate, expecting that the more consonants precede the S the shorter the individual segments, including S, would become. Since this numerical variable had only four values (0, 1, 2 and 3) the variable was

192

transformed into a categorical variable to prevent the model from coming up with nonsensical estimates (e.g. a word with 1.2 consonants in the rhyme). Consonants were counted on the basis of the actual pronunciations.

Frequency. Frequency also affects phonetic duration, with the general tendency of more frequent words exhibiting more reduction, i.e. shorter segment durations (e.g. Bybee 2001: 78; Jurafsky et al. 2001). We used two kinds of frequency, base frequency and form frequency. For base frequency we took the log-transformed frequency of the base in the spoken part of the Corpus of Contemporary American English (COCA, Davies 2008). For words with non-morphemic S we took the frequency of the word. For the form frequency we calculated the log-transformed frequency of the word including the final S, ignoring what kind of S is attached in a given token. Thus, for the words with non-morphemic S, both frequency measures were identical. The frequencies were log-transformed to reduce the potentially harmful effect of skewed distributions in linear regression models.

Neighborhood density. We included phonological neighborhood densities as covariates, as it has been shown that the number of phonological neighbors may influence phonetic reduction (and hence word duration), as denser networks facilitate articulation (see, for example, Gahl et al. 2012). Neighborhood density measures were taken from the CLEARPOND database (Marian et al. 2012). In this database neighbors are those words that differ in one segment from the word in question, with the difference originating from either substitution, addition or deletion of one segment (Marian et al. 2012: 3).

Bigram frequency. Another potential factor influencing the duration of a word in running speech is the predictability of the word in its context (e.g. Jurafsky et al. 2001, Pluymaekers et al. 2005a, Bell et al. 2009, Torreira & Ernestus 2009). For content words, recent studies unanimously show that it is the upcoming context that may have an effect on different aspects of acoustic reduction (including duration, e.g. Pluymaekers et al. 2005a, Bell et al. 2009, Torreira & Ernestus 2009). To account for this durational effect we measured (and log-transformed) the bigram frequency of the item in question and its following word, also based on COCA.

Previous mention. Another reduction effect may arise from online priming. The more often a complex word or its base has been mentioned in the previous discourse, the shorter we expect its duration to be in a given case (e.g. Fowler & Housum 1987, Fowler 1988, Gahl et al. 2012). We therefore included a covariate that counted the number of previous mentions in a time window of 30 seconds preceding the token in question.

Following context. It is known from the literature that the context may have an effect on the duration of consonants. Words and segments (especially fricatives) at the end of an utterance are subject to lengthening (e.g. Oller 1973: 1244; Berkovits 1993). Furthermore, the following segment may have an effect on the duration of the consonant preceding it (e.g. Klatt 1976, Umeda 1977). We therefore coded

193

for each item the segment type of the word following the S as either vowel ('V'), approximant ('APP'), nasal ('N'), affricate ('AFF'), fricative ('F') or plosive ('P'). If S was not followed by any sound, because it occurred at the end of an utterance, we coded the following segment item as '<pause>'. We would expect an S in utterance-final position to be longer than an S we find in other positions.

Syntactic position. Similarly to what happens before pauses, segments before a syntactic or prosodic boundary are also lengthened (see, for example, Klatt 1976, Cooper 1976, Cooper et al. 1978, Byrd et al. 2006). For our data, determining the boundaries of intonation phrases turned out to be highly problematic and unreliable, and we therefore opted for a syntax-based coding. We coded each item for whether it occurred at the right boundary of a syntactic phrase (e.g. an NP). Such a coding can at least partially control for differences that might occur due to syntax-based prosodic effects, for example, between phrase-final plural nouns and pre-head genitives.

3.2.3 *Statistical analysis*

We devised two types of analysis, with different constellations of variables and different statistical models. In the first analysis we used ABSOLUTE DURATION of S as the dependent variable (Model 1), whereas the second analysis had PROPORTION OF S as its dependent variable (Model 2).

Model 1 was fitted using mixed-effects regression, as implemented in the packages `lme4` (Bates et al. 2014) and `lmerTest` (Kuznetsova et al. 2014) for R (R Development coreteam 2011). Mixed-effects regression brings the variation of random effects such as subject or item under statistical control, and can deal with unbalanced data sets. The latter property is especially welcome since not all combinations of all values of the different predictors are represented in our data with equal frequency.

The mixed-effects model was fitted adhering to the following strategy. In the initial model, alongside the explanatory variable TYPE OF S, we included the control variables discussed in the previous section. This initial model was then reduced through step-wise exclusion of insignificant factors (e.g. Baayen 2008). A factor was only considered significant if it passed three tests. First, its *t*-statistics had to yield a *t*-value greater than 2 (or less than −2) when included in the model. Second, the Akaike information criterion (AIC) of the model including the factor had to be lower than the AIC of the model without it. Third, an ANOVA comparing the model including the factor to a model without it had to yield a *p*-value lower than 0.05, thus showing that the inclusion of the factor did significantly improve the fit of the model. A variable under consideration was only retained in the model if it passed all three tests.

One of the central assumptions of any linear regression model is a linear relationship between the dependent and independent variables. If this assumption is not met, the estimated coefficients may be highly unreliable. In this case, the dependent variable may often be transformed to alleviate any problem resulting

from a lack of linearity. The Box–Cox transformation (Box & Cox 1964, Venables & Ripley 2002) can be employed to identify a suitable transformation parameter λ for a power transformation. For the transformation of the absolute duration measures used in Model 1 the optimal value of λ was $\lambda = 0.1010101$.

We also tested interactions between each covariate and the type of S. In addition, we tested an interaction between voicing and the number of consonants in the rhyme of the base-final syllable, since the presence of more consonants increases the distance between the S and the voiced nucleus of the syllable, which might lead to less assimilation.

Let us turn to Model 2. In this analysis the dependent variable was the relative length of S (PROPORTION OF S, calculated as the (non-transformed) absolute duration of the S divided by the duration of the whole word). As this variable is bounded between 0 and 1 and has a skewed distribution, linear regression is ruled out and a model is needed that can cope with these properties of the target variable. Beta regression (e.g. Ferrari & Cribari-Neto 2004) is such a model. We used the R package betareg (Cribari-Neto & Zeileis 2010) for this analysis. For the beta regression models a slightly different fitting strategy had to be adopted. This will be explained in more detail in Section 4.2.

Collinearity is an issue for two pairs of covariates, the first of which is the number of syllables and the duration of the base, the second the two frequency measures base frequency and form frequency. Standardly used residualization procedures have recently been shown to be inadequate tools to address these issues (e.g. Wurm & Fisicaro 2014). We therefore adopted a different strategy. The number of syllables and the duration of the base were highly correlated and inclusion of both into our models led to suppression effects (e.g. Holling 1983). Since the direction of influence on the duration of S was the same for both covariates, and models with only one of the two covariates showed that the influence of duration of the base on the duration of S was much stronger, we only included the duration of the base into our analyses. Similarly, base frequency and form frequency were highly correlated and showed the same direction in their effect on the duration of S when looked at individually. The effect of form frequency was much weaker, however. We therefore included only base frequency into our models.

3.2.4 *Overview of the data*

An overview of variables and their distributions is given in Table 2. The names of the (sometimes transformed) variables that entered the analysis are given in small capitals. Note that not all variables were available for all observed morphemes: some bases are not listed in CLEARPOND (e.g. *permeate* or *sandal*), so no neighborhood density information could be used for these observations. Also, LOGRBIGRAM was not calculated for observations that were followed by a speech pause in the recording.

195

Dependent variables	N	Mean	St. Dev.	Min	Max
Absolute duration of S: DURATIONOFS	644	0.081	0.039	0.019	0.237
Relative duration of S: PROPORTIONS	644	0.206	0.082	0.055	0.688

Numerical predictors	N	Mean	St. Dev.	Min	Max
Local speech rate: SYLSEC	644	5.601	1.202	1.984	10.179
Base duration: BASEDURATION	644	0.329	0.134	0.029	1.052
Base frequency: LOGBASEFREQ	644	8.672	2.399	0.000	14.146
Previous mention: BASEREP	644	0.317	0.772	0	6
Bigram frequency: LOGRBIGRAM	548	2.542	2.739	0.000	9.884
Neighborhood density: PND	601	14.218	14.855	0	60

Categorical predictors:	N	Levels			
No. of cons. before S: CONSONANTS	644	0: 325	1: 259	2: 58	3: 2
Voicing: ISVOICED	644	voiced: 81	unvoiced: 563		
Following context: FOLLCONTEXT	644	\<pause\>: 97	V: 170	APP: 68	N: 33
		AFF: 10	F: 143	P: 123	
Syntactic position: BOUNDARY	644	yes: 226	no: 418		

Explanatory variable	N	Levels			
Type of S: TYPEOFS	644	s: 196	PL: 95	3rdsg: 100	GEN: 88
		has: 47	is: 95	PL-GEN: 23	

Table 2
Summary of the dependent variables and covariates used in the initial models.

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr (> F)
TYPEOFS	0.01	0.00	6.00	521.35	2.76	0.0121
ISVOICED	0.01	0.01	1.00	603.91	24.48	0.0000
CONSONANTS	0.02	0.01	3.00	590.04	14.20	0.0000
FOLLCONTEXT	0.13	0.02	6.00	599.24	51.92	0.0000
SYLSEC	0.00	0.00	1.00	604.56	10.35	0.0014
BASEDURATION	0.01	0.01	1.00	606.83	23.62	0.0000
TYPEOFS:ISVOICED	0.01	0.00	6.00	596.98	3.63	0.0015

Table 3
p-values of fixed effects in Model 1, fitted to the Box–Cox-transformed durations of S.

4. RESULTS

4.1 *Model 1: Absolute duration as dependent variable*

Model 1 was fitted according to the procedure described above. The inspection of the residuals showed a non-normal distribution in both tails of the distribution. Following standard procedures (e.g. Crawley 2002, Baayen & Milin 2010), we removed outliers (defined as items with standardized residuals exceeding −2.5 or +2.5) and refitted the model (the removal of outliers resulted in the loss of 1.7% of the observations). The final model showed a satisfactory distribution of residuals.

In the final model we find significant main effects of type of S (TYPEOFS), the following context (FOLLCONTEXT), the number of consonants preceding S (CON-SONANTS), speech rate (SYLSEC) and duration of the base (BASEDURATION). In addition, there is a significant interaction of type of S (TYPEOFS) and voicing (ISVOICED). Regarding the random effects, we tested speaker and base, but only speaker-specific effects turned out to significantly improve model performance. The *p*-values for the analysis of variance (or deviance) of Model 1 are documented in Table 3.

The *R*-squared value of the model is 0.65, with the random effect explaining 6 percent of the variation, and the fixed effects 59 percent. We checked the effect sizes of the individual predictors by fitting models that lacked a particular predictor, and compared their *R*-squared values with the *R*-squared value of the model that contained all six predictors. The hierarchy in (4a) reflects the results. The decrease in the *R*-squared value is greatest when taking out FOLLCONTEXT, followed by ISVOICED, followed by TYPEOFS, and so forth. Furthermore, we devised ANOVAs to check whether a model that lacked a given predictor performed better than a model that lacked a different predictor. This resulted in the hierarchy given in (4b), which gives the most powerful predictors on the left. The models that lack one of either FOLLCONTEXT, TYPEOFS, ISVOICED do not show a significant difference in pair-wise ANOVAs, but perform significantly worse than models that lack one of the variables to their right. It should be noted, however, that taking out either ISVOICED or TYPEOFS also eliminates the

197

interaction term from the model, such that we are actually losing two predictors in those cases. We therefore also fitted models with only one of the six variables. These models resulted in the hierarchy in (4c), which reflects decreasing *R*-squared values from left to right. Overall, the morphological status of an S thus turns out to be a rather strong predictor of its acoustic duration.

(4) (a) FOLLCONTEXT >> ISVOICED >> TYPEOFS >> CONSONANTS >> BASEDURATION >> SYLSEC

  (b) FOLLCONTEXT, ISVOICED, TYPEOFS >> CONSONANTS >> BASE-DURATION >> SYLSEC

  (c) FOLLCONTEXT >> TYPEOFS >> BASEDURATION >> CONSO-NANTS >> ISVOICED,SYLSEC

The estimates (and their *p*-values) of Model 1 are documented in Table 4. The reference levels for the categorial predictors are the following: for TYPEOFS it is non-morphemic S, for FOLLCONTEXT it is <pause>, for ISVOICED it is voiced and for CONSONANTS it is 0. All coefficents can be interpreted as changes relative to these reference levels.

Figure 3 illustrates the effects of the non-interacting covariates in Model 1. The estimated values of the dependent variable are back-transformed into seconds. Three of the four covariates that come out as significant show the expected behavior. In the upper row the left panel (FOLLCONTEXT) shows that before a pause, i.e. in utterance-final position, S is longer than in other positions where there is a segment following S. This can be interpreted as a clear final-lengthening effect. In comparison, the different types of following segments shown in the panel have a relatively small effect on the duration of S. Apparently, the strong effect size of FOLLCONTEXT seen in (4) above is almost completely attributable to positional lengthening. The right panel of the upper row shows the effect of the number of consonants before S. The more consonants we find in the rhyme the shorter final S becomes, which is equally expectable.[6] The left lower panel gives us speech rate (SYLSEC). With faster speech, S becomes shorter, as predicted. In the right panel of the lower row we see the effect of BASEDURATION. Words that have a longer duration also have longer S's, which also means that final S participates in the overall lengthening or shortening effects that affect the word as such. If we conceive of base duration as a measure of very local speech rate this is expected: if the very local speech tempo is low, the base duration is long, and one would expect the S to be long also. If, however, we expect a kind of compression effect (based on what was found for vowels in Dutch), this effect is unexpected and unclear in its interpretation. We fitted two separate models for monosyllables and disyllables, respectively, to hold syllable number constant. The same effect of base duration emerged in both of these separate models, which corroborates

---

[6] The effect trails off for three consonants, but we have only two observations for this value and cannot say much about this value of CONSONANTS.

	Estimate	Std. Error	df	*t*-value	Pr (> |*t*|)
(Intercept)	0.795	0.009	607.947	91.979	0.000
TYPEOFSplural	0.018	0.010	605.021	1.761	0.079
TYPEOFS3rdsg	−0.014	0.010	603.634	−1.394	0.164
TYPEOFSGEN	0.005	0.009	606.894	0.638	0.524
TYPEOFShas	0.002	0.009	604.082	0.242	0.809
TYPEOFSis	−0.001	0.008	607.858	−0.126	0.900
TYPEOFSPL-GEN	0.020	0.014	604.843	1.426	0.154
ISVOICEDunvoiced	0.034	0.006	604.528	5.371	0.000
CONSONANTS1	−0.008	0.002	601.617	−3.747	0.000
CONSONANTS2	−0.021	0.003	593.097	−6.370	0.000
CONSONANTS3	−0.019	0.015	577.108	−1.271	0.204
FOLLCONTEXTAFF	−0.043	0.007	603.678	−5.933	0.000
FOLLCONTEXTAPP	−0.038	0.004	602.581	−10.327	0.000
FOLLCONTEXTF	−0.045	0.003	604.112	−14.946	0.000
FOLLCONTEXTN	−0.043	0.004	592.280	−9.685	0.000
FOLLCONTEXTP	−0.052	0.003	600.168	−16.092	0.000
FOLLCONTEXTV	−0.041	0.003	599.933	−14.028	0.000
SYLSEC	−0.003	0.001	604.556	−3.217	0.001
baseDuration	0.036	0.007	606.829	4.860	0.000
TYPEOFSplural:ISVOICEDunvoiced	−0.030	0.010	591.908	−2.987	0.003
TYPEOFS3rdsg:ISVOICEDunvoiced	0.003	0.010	595.894	0.268	0.789
TYPEOFSGEN:ISVOICEDunvoiced	−0.024	0.009	604.859	−2.627	0.009
TYPEOFShas:ISVOICEDunvoiced	−0.025	0.010	604.523	−2.545	0.011
TYPEOFSis:ISVOICEDunvoiced	−0.021	0.008	602.244	−2.536	0.011
TYPEOFSPL-GEN:ISVOICEDunvoiced	−0.037	0.015	606.678	−2.464	0.014

Table 4
Fixed-effect coefficients and *p*-values in Model 1 (mixed-effects model fitted to the Box–Cox-transformed durations of S).

the idea that we find no compression effect, but S participating in lengthening or shortening effects of the word it occurs in.

Let us now turn to the variable of interest, i.e. TYPE OF S, whose effect is plotted in Figure 4 in interaction with voicing. Again, the values of the dependent variable are back-transformed into seconds.

We can see that, apart from genitive-plural, voiced S is generally shorter than unvoiced S. This general trend is as expected. What is less expected is that the degree to which voiced realizations differ in duration from unvoiced realizations is dependent on the type of S. Non-morphemic and 3rd person singular S show a large difference in duration between voiced and unvoiced realizations, while plural, genitive and the two clitics show markedly smaller differences.
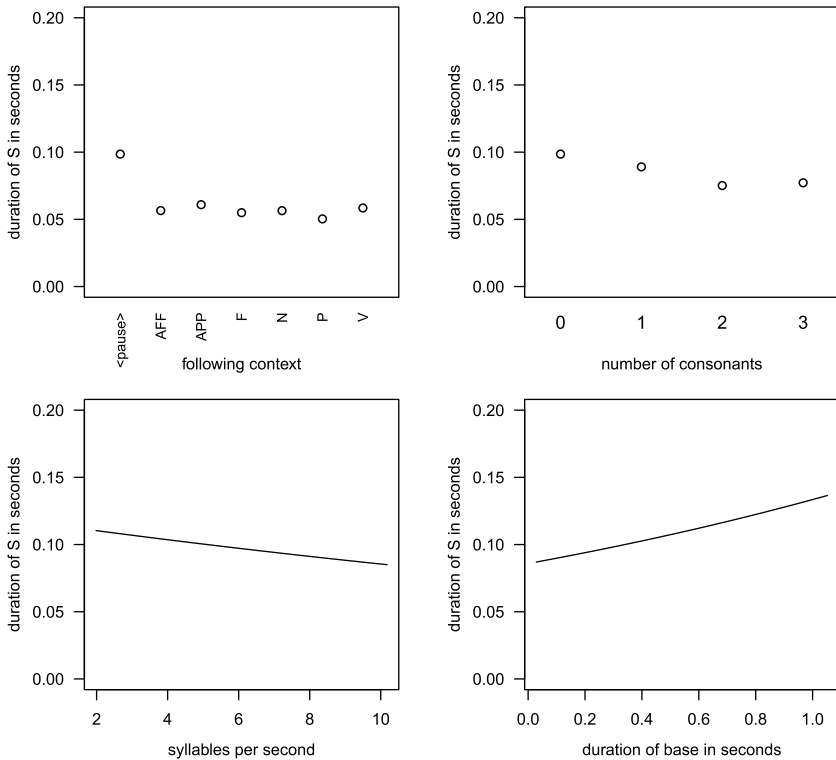
199

Figure 3
Partial effects of non-interacting covariates in Model 1, fitted to the Box–Cox-transformed
absolute durations of S.

Testing of all pair-wise contrasts between the different types of S yields the significant contrasts shown in Tables 5 and 6. We compare the different types of S while holding voicing constant.

If we compare the two tables, we can first see that we find fewer contrasts for voiced realizations of S. This is not surprising for two reasons. First, this subset is much smaller, with 81 voiced as against 552 unvoiced items. Second, voiced items are generally shorter, which makes it harder to detect differences. However, even in this subset we find four significant contrasts. Third person singular S is significantly shorter than plural, genitive and genitive-plural. In addition, plural S is significantly longer than the *is* clitic.

For unvoiced S, 10 of the overall 21 possible pair-wise contrasts are significant. Non-morphemic S is significantly longer than all types of morphemic S. The other most remarkable pattern concerns the two clitics of *has* and *is*. They are significantly shorter than 3rd person singular S and plural S.
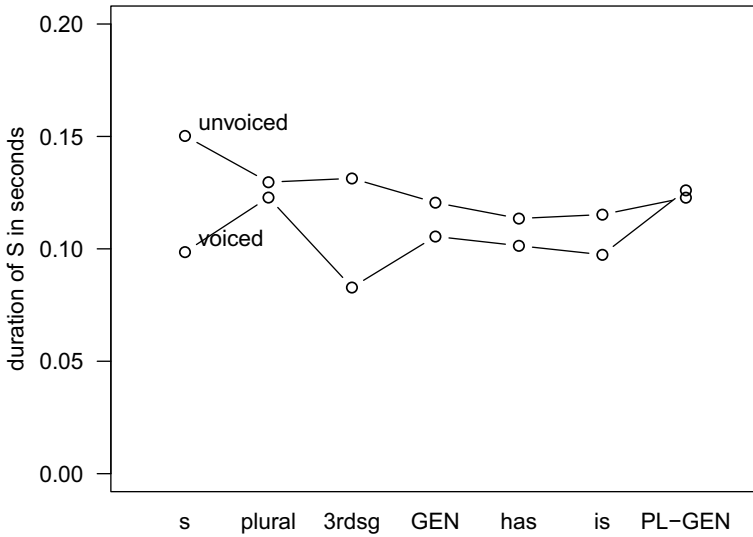
200

Figure 4
Interaction of type of S and voicing, Model 1 (Abbreviations: s = non-morphemic S, 3rdsg = 3rd person singular, GEN = genitive, PL-GEN = genitive-plural).

	S	PL	3RDSG	GEN	HAS	IS	PL-GEN
S	n.a.						
PL		n.a.	*			*	
3RDSG			n.a.	*			*
GEN				n.a.			
HAS					n.a.		
IS						n.a.	
PL-GEN							n.a.

Table 5
Significant contrasts in duration between different types of voiced S. Significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

The results from Model 1 clearly indicate that the null hypotheses in (1) and (2) need to be rejected. We find robust differences between morphemic and non-morphemic S with unvoiced S. We also find robust differences between some types of morphemic S with both voiced and unvoiced S. These differences are present in natural conversational speech, and they cannot be attributed to purely phonetic or lexical effects since these effects were carefully controlled for.

201

	S	PL	3RDSG	GEN	HAS	IS	PL-GEN
S	n.a.	**	*	***	***	***	**
PL		n.a.			*	*	
3RDSG			n.a.		*	*	
GEN				n.a.			
HAS					n.a.		
IS						n.a.	
PL-GEN							n.a.

Table 6
Significant contrasts in duration between different types of unvoiced S. Significance codes:
'***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

4.2 *Model 2: Relative duration*

An inspection of the distribution of the relative duration measurements showed three outliers where the S was extremely long compared with the rest of the items, with proportions being larger than 54 percent. These three items were removed. The distribution of the measurements of relative length of S was substantially skewed, which is something that can be frequently observed with ratios that are bounded between 0 and 1. Beta regression is a kind of statistical model that can cope with these distributional properties. Beta regression models can contain two components, one that predicts the mean and a component for the precision phi (see Ferrari & Cribari-Neto 2004 for introduction and discussion). Broadly speaking, a predictor with a low precision coefficient means that the beta regression model estimates the values of this predictor to be more dispersed around the coefficient's mean than in the case of a predictor with a high precision coefficent.

We fitted a beta regression model with the proportion of S as the dependent variable, and the same initial fixed effects as in Model 1 apart from BASEDU-RATION, since this variable had been used in the computation of the dependent variable. For the beta regression models a slightly different fitting strategy from that of mixed effects regression had to be adopted (see Ferrari & Cribari-Neto 2004 for details). The initial models started with only the mean component and all predictors. The initial model was then simplified in a step-wise procedure, removing all insignificant predictors. After the completion of simplification for the mean component, we added the precision component and then simplified this component applying again the standard step-wise procedures. After completion of this simplification procedure, we finally tested whether the inclusion of the precision component as a whole was justified by comparing the AIC of a model that had only the mean component with the model with both mean and precision components, and by using a likelihood ratio test. The inclusion of the precision parameters was justified by lower AIC scores and higher log-likelihood. The estimated coefficients for the means and the phi-values of the final model are

202

given in Table 7. The reference levels are the same as with Model 1, so that all coefficients can be interpreted as changes relative to these reference levels. The bottom part of Table 7 documents the precision model, which outputs a significant effect for TYPE OF S, voicing and following context.

With regard to our explanatory variable, Model 2 tells us that both the means and the dispersion of the relative length of S are dependent on which kind of S we look at. For the means, the effect of type of S interacts with voicing. This is fully in line with the results from Model 1.

The behavior of the covariates in the mean model is very similar to their behavior in the model that tried to predict the absolute duration of S. This is illustrated in Figure 5.[7] Each panel shows the partial effect of a different covariate, with the values of the covariate given on the horizontal axis. The vertical axis corresponds to the estimated relative duration of S. The dots indicate the estimated means, and the bulb-shaped areas indicate the estimated beta density for the values of the covariates. The shapes of these areas inform us about the estimated dispersion around the mean: short areas (e.g. for fricatives in the upper-left panel) indicate that the model estimates the variance around the mean to be relatively low for this type of following context, while elongated areas (e.g. approximants in the same panel) suggest a relatively high variance.

The main effects of following context, number of consonants, base duration and number of syllables per second are found in both Model 1 and Model 2 and they go in the same direction as before. In middle position the proportion of S is smaller (see the negative coefficient for all segments), which means that we find the expected effect of final lengthening. This is in line with the results of pertinent investigations (e.g. Turk & Shattuck-Hufnagel 2007). With regard to rhyme structure, the more consonants there are in the rhyme the smaller the proportion of S becomes (see the negative coefficients), which is again an expectable effect. Higher speech rate leads to shorter realizations of S, and S participates in lengthening or shortening effects of the whole word.

Let us turn to the variable of interest, type of S. Figure 6 shows the effect of the interaction between type of S and voicing on the estimated relative durations, and Table 8 gives the estimated means in unvoiced and voiced realizations for the different values of TYPE OF S.

An area surrounded by a solid line indicates the beta density for the voiced realizations of the associated morpheme type. The corresponding dashed area indicates the density for the unvoiced realizations. The filled and circled dots indicate the mean relative duration for voiced and unvoiced realizations (see Table 8). Note that the beta density is not shown for voiced genitive-plural. For this factor combination, the estimated precision coefficient phi exceeds the valid range for which the beta density can still be calculated. Thus, the variance is

---

[7] The R function used to create Figures 5 and 6 is available here: https://bitbucket.org/gkunter/pl otbetareg.

Coefficients (mean model with logit link)

	Estimate	Std. Error	z-value	Pr (> |z|)
(Intercept)	−0.135	0.104	−1.297	0.195
TYPEOFSplural	0.229	0.111	2.063	0.039
TYPEOFS3rdsg	−0.183	0.114	−1.608	0.108
TYPEOFSGEN	−0.059	0.090	−0.651	0.515
TYPEOFShas	−0.048	0.103	−0.465	0.642
TYPEOFSis	−0.102	0.084	−1.214	0.225
TYPEOFSPL-GEN	0.164	0.115	1.422	0.155
ISVOICEunvoiced	0.432	0.072	6.017	0.000
CONSONANTS1	−0.088	0.027	−3.226	0.001
CONSONANTS2	−0.324	0.048	−6.782	0.000
CONSONANTS3	−0.336	0.257	−1.306	0.192
FOLLCONTEXTaffricate	−0.380	0.129	−2.953	0.003
FOLLCONTEXTapproximant	−0.310	0.051	−6.069	0.000
FOLLCONTEXTfricative	−0.490	0.043	−11.485	0.000
FOLLCONTEXTnasal	−0.434	0.062	−6.945	0.000
FOLLCONTEXTplosive	−0.564	0.045	−12.415	0.000
FOLLCONTEXTvowel	−0.407	0.041	−9.986	0.000
SYLSEC	−0.052	0.010	−5.316	0.000
BASEDURATION	−2.327	0.103	−22.598	0.000
TYPEOFSplural:ISVOICEunvoiced	−0.394	0.119	−3.315	0.001
TYPEOFS3rdsg:ISVOICEunvoiced	0.057	0.120	0.474	0.635
TYPEOFSGEN:ISVOICEunvoiced	−0.204	0.098	−2.088	0.037
TYPEOFShas:ISVOICEunvoiced	−0.270	0.114	−2.367	0.018
TYPEOFSis:ISVOICEunvoiced	−0.178	0.092	−1.934	0.053
TYPEOFSPL-GEN:ISVOICEunvoiced	−0.450	0.126	−3.576	0.000

Phi coefficients (precision model with log link)

	Estimate	Std. Error	z-value	Pr (> |z|)
(Intercept)	4.457	0.238	18.741	0.000
TYPEOFSplural	−0.256	0.176	−1.452	0.146
TYPEOFS3rdsg	−0.244	0.174	−1.401	0.161
TYPEOFSGEN	0.367	0.186	1.978	0.048
TYPEOFShas	0.200	0.252	0.793	0.428
TYPEOFSis	0.448	0.184	2.440	0.015
TYPEOFSPL-GEN	0.875	0.315	2.778	0.005
ISVOICEunvoiced	−0.591	0.177	−3.351	0.001
FOLLCONTEXTaffricate	−0.536	0.469	−1.143	0.253
FOLLCONTEXTapproximant	0.051	0.227	0.224	0.823
FOLLCONTEXTfricative	0.361	0.189	1.911	0.056
FOLLCONTEXTnasal	0.164	0.290	0.563	0.573
FOLLCONTEXTplosive	0.357	0.204	1.748	0.081
FOLLCONTEXTvowel	0.486	0.182	2.670	0.008

Table 7
Coefficients of final beta regression model, Model 2 (pseudo-R^2 of the model: 0.596).
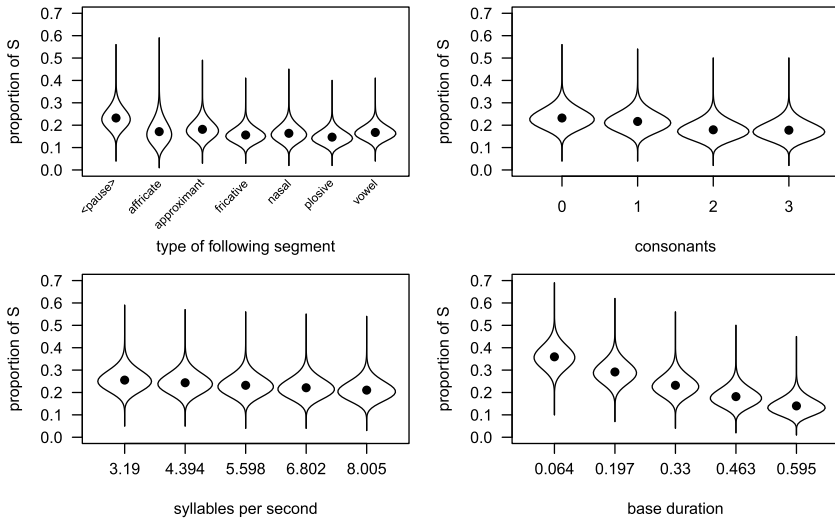
Figure 5
The effect of the non-interacting covariates on the relative duration of S, final beta
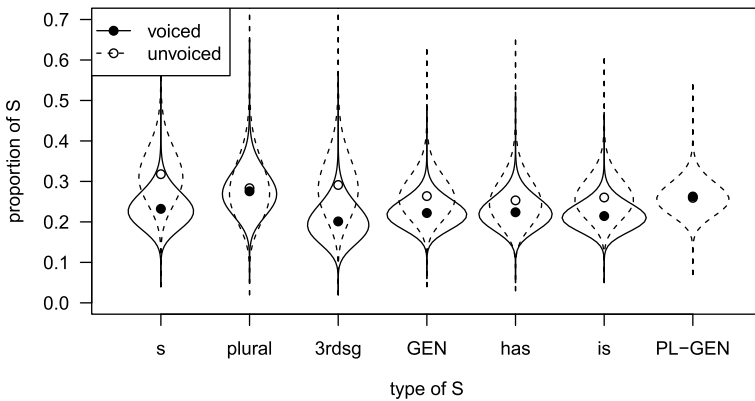regression model, Model 2.

Figure 6
Effect of the interaction between morpheme type and voicing on the relative duration of S,
final beta regression model, Model 2. (Abbreviations: s = non-morphemic S, 3rdsg = 3rd
person singular, GEN = genitive, PL-GEN = genitive-plural).

estimated to be virtually minimal – given that this estimation is based on only
three observations, this result is not very surprising.

For the majority of types of S, the model estimates the mean relative duration
to be different for voiced and unvoiced realizations. Investigation of the contrasts
reveals that the differences for non-morphemic S, 3rd person singular, genitive

205

Type of S	voiced	unvoiced
non-morphemic S	0.232	0.318
plural	0.276	0.283
3rdsg	0.201	0.291
genitive	0.222	0.264
has	0.224	0.253
is	0.215	0.261
genitive-plural	0.263	0.260

Table 8
Predicted means of the different types of S for voiced and unvoiced realizations, final beta
regression model, Model 2.

	S	Plural	3rdsg	GEN	has	is	PL-GEN
S	n.a.	*					
plural		n.a.	**	**	*	*	
3rdsg			n.a.				**
GEN				n.a.			*
has					n.a.		
is						n.a.	*
PL-GEN							n.a.

Significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

Table 9
Significant contrasts in relative duration between different types of voiced S.

and *is* are statistically significant (the difference for *has* is only marginally signif-
icant). Much more interesting with regard to our null hypothesis are, however, the
differences between the different types of S. If we pair-wise compare the different
estimates in the mean component of Model 2, we find eight significant contrasts
for voiced realizations, and 12 (out of 21 possible ones) for unvoiced realizations.
For these pairs, the beta regression estimates that the means of the relative duration
differ significantly from each other. Tables 9 and 10 summarize the contrasts.

The pair-wise contrasts for the relative duration of the different types of S
yield an interesting picture. For voiced realizations, plural is significantly longer
than all other types of S, except for genitive-plural. Genitive-plural, in turn, is
significantly longer than 3rd singular S and the *is* clitic. For unvoiced realizations,
a different pattern emerges. Non-morphemic S is significantly longer than all other
types of S. The two suffixes plural and 3rd person singular have similar relative
durations, but they are significantly longer than the two auxiliary clitics *has* and
is. In addition, 3rd person singular is also significantly longer than genitive and
genitive-plural (plural shows a marginally significant contrast to genitive and

206

	s	Plural	3rdsg	GEN	has	is	PL-GEN
s	n.a.	***	**	***	***	***	***
plural		n.a.			*	*	
3rdsg			n.a.	**	**	**	**
GEN				n.a.			
has					n.a.		
is						n.a.	
PL-GEN							n.a.

Significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

Table 10
Significant contrasts in relative duration between different types of unvoiced S.

genitive-plural, $p = 0.06$ for both of these contrasts). The relative durations of the four clitics (i.e. *has*, *is*, genitive and genitive-plural, if the latter is analyzed as clitic) do not differ significantly from each other. In terms of relative duration of unvoiced realization, we can thus determine three groups: non-morphemic S, the suffixes and the clitics (the latter including genitive-plural).

To summarize, the analysis of the relative duration of S has shown that there are many significant differences between different types of S, in particular between plural and the rest for voiced realizations, and between non-morphemic, suffixal and clitic S for unvoiced realizations. The results of Model 2 thus support and complement the results of Model 1, to the effect that the null hypotheses in (1) and (2) need to be rejected for both kinds of duration measurements.

5. SUMMARY AND DISCUSSION

5.1 *Summary of results*

Following in the footsteps of recent studies of free lexemes by, for example, Drager (2011) and Gahl (2008), we tested whether the morphological status of S has an influence on the acoustic duration of this segment. We started out from two null hypotheses. The first stated that there are no durational differences between non-morphemic S and morphemic types of S. The second null hypothesis stated that there are no durational differences between the different homophonous types of morphemic S. We investigated the duration of the different types of S using conversational data. No matter whether we looked at the absolute duration of S or the relative duration of S, the type of S emerged as a strong, significant predictor of segmental duration. We found statistically significant differences between certain types of S. This means that we have to reject the two null hypotheses.

The details of the results are intricate, but the clearest outcomes can be found with unvoiced realizations. For this subset, the relative duration shows a pattern of differences that correlates with the type of morphological boundary.

207

Words with no morphological boundary before S show the longest duration of S. Words with a suffixal S have shorter S's, and cliticized S's are even shorter (the difference between plural and genitive, and plural and genitive-plural was marginally significant). The results for voiced realizations are very similar and show a difference between the two suffixes and the two clitics. Even though the voiced realizations show fewer and somewhat different contrasts, they still clearly falsify both null hypotheses.

Our analyses of absolute and relative duration of the different types of S included pertinent covariates that controlled for purely phonetic effects known from the phonetic literature. The covariates whose effects are well established by other studies behave as expected, so that one can assume that the speech corpus data used for this study are generally reliable. Thus, the results, surprising though they may be, cannot be easily dismissed as due to an unsuitable sample, or an inadequate phonetic analysis.

5.2 *Comparison of results to other studies*

How do our results relate to those of previous studies? Let us first discuss the difference between morphemic S and non-morphemic S (Null Hypothesis 1, see (1)). Our results concerning the absolute duration of S seem to agree with Walsh & Parker (1983), who also found differences between the two types of S. Such a statement would, however, be premature. As these authors did not employ any inferential statistics, or consider any covariates in their analysis, we cannot determine whether the small differences found in the means by Walsh & Parker are meaningful. It is noteworthy, however, that the mean difference observed by Walsh & Parker goes in the opposite direction from ours: in their data, morphemic S is longer, not shorter, as it is in our sample. However, their mean difference between the morphemic (i.e. plural in their experiment) and the non-morphemic /s/'s was only 6 ms, while we observed much larger differences in the means (e.g. 38 ms estimated difference between non-morphemic S and the *has* clitic, and 47 ms difference in the observed means, see Table 1).

In a more recent study that investigated nine monosyllabic word types ending in /z/, Song et al. (2013) found a significant difference between plural and non-morphemic /z/ in utterance-final position, but not in non-final position. Morphemic /z/ was longer than non-morphemic /z/, but only by 7 ms.[8]

It is not clear where the discrepancies between those data and ours come from, but we may speculate about possible reasons. One reason might be the different kinds of item. In Walsh & Parker's experimental study, morphologically complex words were compared with their simplex homophones, while in our data we tested the duration of S across the board, i.e. across many different, non-homophonous, types. It may well be that phonetic differences between

---

[8] This is the overall mean across positions. Unfortunately, the paper does not give the means for utterance-final position.

phonologically homophonous word forms of two different lexemes arise from the competition of these lexemes (or their word forms) in speech production.

Another difference between Walsh & Parker's and the present study is the nature of the data. The reading out loud of the stimuli list may have influenced the results in a way that we do not find in spontaneous speech. Gahl et al. (2012), for example, discuss studies that have failed to reproduce the well-established effect that more frequent words generally show shorter duration. The authors remark that this failure to replicate the effect may be due to the fact that speakers tend to read at a regular pace when asked to read word lists or words in short carrier phrases. This tendency has been found to override effects of lexical properties, such as word frequency.

Song et al.'s study is based on conversational speech, i.e. data that seem comparable to ours. However, their data set is restricted to monosyllables and to only nine different word types. Furthermore, the set of covariates in that study was small and potential variability in voicing was not taken into account when analyzing the duration of morphemic versus non-morphemic /z/. Another, and perhaps important, difference from our study is that Song et al.'s data comprise child-directed speech, which has been shown to differ from inter-adult speech in various ways (see, for example, Foulkes et al. 2005 for an overview and discussion). More research is certainly called for to replicate the effects observed in previous studies and the present one.

With respect to differences between different types of morphemic S, there are hardly any studies available with which we could compare our results. Song et al. (2013) do not find a difference between 3rd singular and plural S, but they limited their data set to only three word types for each morpheme. Li et al. (1999), also investigating mothers' speech to their young children, found that plural S was longer than 3rd person singular S, but attributed this difference to the fact that plural nouns were more prone to appear in utterance-final position in their data set. In our data, we also found an effect of position but no interaction of position with type of S. The lack of studies and the contradictory results of the few existing studies call for more research into the phonetic realization of morphemic S in English.

5.3 *Explanations and implications*

The results of the present study raise important theoretical questions. Most importantly, it is unclear why the differences between different types of S that we find in our data would emerge in the first place. More specifically, one is tempted to ask the question why a certain type of S should be longer than some other S. For instance, why are the clitics particularly short and the suffixes rather long (and non-morphemic S even longer)? Which theory would predict such results?
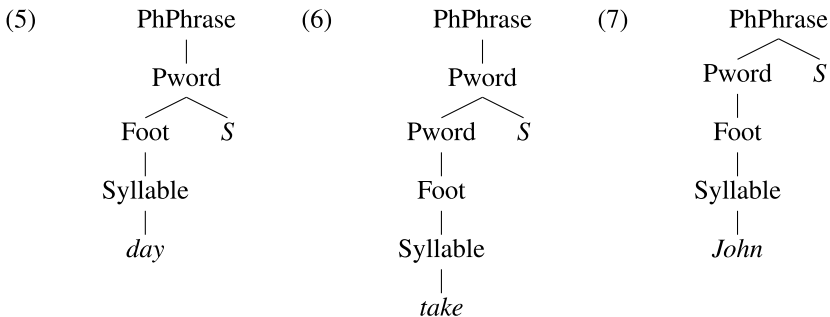
At a very general level, our findings can be interpreted as support for the idea that there is morphological information in the phonetic signal, i.e. in post-lexical stages of speech production. This calls into question the distinction

between lexical and post-lexical phonology, which has featured prominently in both theoretical linguistics, i.e. phonology, and psycholinguistics.

In phonology, there is a long tradition of theoretical mechanisms like bracket erasure and cyclic application of morpho-phonological rules, after whose application there is no possibility that one could trace any information about a sound's origin or structural status in the acoustic signal. The findings of this paper thus seriously challenge central tenets of the traditional models of lexical phonology and morphology in the wake of Kiparsky (1982) (see also, for example, Scheer 2010 and Coetzee & Pater 2011 for critical discussion).

There are, however, some alternative explanations conceivable. We will first discuss a prosodic approach and then turn to psycholinguistic ones. In a prosodic phonology approach phonetic differences (for example in duration) between different types of S might emerge from different positions of these S's in the prosodic structure. It has been claimed (e.g. Goad 1998, Goad et al. 2003) that plural S is an 'internal clitic', i.e. it is part of the prosodic word, as in (5), whereas 3rd singular S is an 'affixal clitic' adjoined to the prosodic word formed by its base, see (6). The auxiliary and genitive clitics would be analyzed as 'free clitics', given in (7).

(5)
```
        PhPhrase
           |
         Pword
         /   \
      Foot    S
        |
     Syllable
        |
       day
```

(6)
```
        PhPhrase
           |
         Pword
         /   \
      Pword   S
        |
      Foot
        |
     Syllable
        |
       take
```

(7)
```
        PhPhrase
         /    \
      Pword    S
        |
      Foot
        |
     Syllable
        |
       John
```

This prosodic analysis thus posits a difference between free clitics and affixal clitics that would mirror the difference in duration between the clitics and the suffixes in our study: affixal clitics are shorter than free clitics. However, that would possibly predict that plural should be shorter than 3rd singular. This is, however, not the case. In unvoiced realizations there is no difference between plural and 3rd singular, and in voiced positions there is a difference but it goes in the opposite direction, i.e. plural is longer. In other words, the increasing degree of integration posited from free clitic to internal clitic is not mirrored by a consistent patterning of the acoustic correlate we measured.

In addition, it is not clear why the phonetic properties would pattern in the way they do. A prosodic account has no obvious explanation for why less prosodic integration would go together with shorter duration. On top of that it seems mysterious that voiced and unvoiced realizations can show different kinds of contrasts, although the prosodic structure remains the same for both voiced and unvoiced realizations.

An alternative to a prosodic account is exemplar-based models (e.g. Goldinger 1998; Bybee 2001; Pierrehumbert 2001, 2002; Johnson 2004; Gahl & Yu 2006), which seem to be better equipped to deal with the kind of variation we find in our data. In such models a given word is linked to a frequency distribution over phonetic outcomes, as encountered in the environment of the speaker. These distributions are updated with new experiences, and subtle subphonemic differences in these experiences may result in representations that reflect these properties. For example, Pierrehumbert (2002) demonstrates how such a model can deal with the phonetic variability of lexemes. It is conceivable that similar models can be implemented to account for the subtle phonetic properties of different bound morphemes. The details of such an account still need to be worked out in future studies, however, since available exemplar-based approaches have not included the subtle phonetic differences involved in the differentiation of allegedly homophonous affixes. One problem to solve in an exemplar model is the effects of the covariates. It is not obvious how these general effects could be derived in a pure, non-abstracting model (see also Pisoni & Levi 2009 for discussion).

Turning to psycholinguistic models, well-established models of speech production and the mental lexicon seem equally unable to accommodate our findings. Levelt et al. (1999), for example, assume that phonological representations are composed of discrete segments and syllables, and the articulator module makes use of pre-programmed gestures that are stored in a syllabary (Levelt et al. 1999: 5). The articulator cannot provide a pre-programmed gesture for each syllable of a language if different meanings cause differences in these gestures. In other words, in such models morphologically dependent subphonemic detail is not part of these representations and needs therefore to be accounted for by purely phonetic factors that influence articulatory implementation such as speech rate (e.g. Levelt 1989). For our data, such an account is ruled out.

There is also a line of psycholinguistic research that has looked at distinct processing properties of different kinds of morphemes. Using data from aphasia, second language acqusition and code switching, Myers-Scotton & Jake (2000) propose a four-way distinction between different types of morpheme (the so-called '4-M model'). According to the 4-M model there are content morphemes and system morphemes, and the system morphemes are further subdivided into 'early system morphemes' and 'late system morphemes'. These classes are not all very well defined, but it seems clear that plural belongs to the early system morphemes while genitive and 3rd person singular marking belong to the late system morphemes. The late system morphemes further consist of two subclasses, 'bridge system morphemes' and 'outsider system morphemes'. According to Myers-Scotton & Jake (2000), genitive *S* is a bridge morpheme, while 3rd person singular is an outsider morpheme. The proposed four-way classification of morphemes is supported by the differential behavior of these morphemes in aphasia, second language acquisition and code switching, and these authors relate the differences in behavior to differences in lexical access. For content and early

211

system morphemes, lexical access happens at the lemma level, and for the late system morphemes at what they call the 'functional level', which is located at the level of the formulator (see, for example, Levelt 1989).

Unfortunately, it is not obvious how this model would be able to account for our results. Even if we assume that Myers-Scotton & Jake's taxonomy is on the right track, there are severe problems. The first one is that their model does not say anything about the details of phonetic implementation, and it is unclear what the model would predict for this stage of speech production. If the 4-M model carries over to articulation (in ways still to be investigated), we would expect plural (being an early system morpheme) to pattern with the auxiliary clitics, which is obviously not the case. What is more, the 4-M model seems to rest on the assumption that the morphemes are accessed separately from their bases. More recent research in morphological processing has cast serious doubt on this idea, and the jury is still out on the question of the psycholinguistic status of individual inflectional morphemes.

To summarize, we have to state that both phonological theory and extant psycholinguistic models fail to provide a convincing explanation of the kind of morphologically-induced phonetic variation that we find in our data.

5.4 *Directions for future research*

As became obvious in the preceding paragraphs, more research is needed to address the many questions the present study raises. First, there is a need for investigations that replicate the effects reported in this paper. These should be both experimental (in order to specifically test the two null hypotheses introduced above) and based on data from other corpora of conversational speech (in order to be directly comparable to the present results that were based on the Buckeye corpus).

Another study that suggests itself given the present findings on S is to look at other homophonous affixes, inflectional -*ed*, i.e. /t/ and /d/ in particular. Based on the present study the hypothesis could be tested that non-morphemic D would be longer than morphemic D, and that suffix D (as in past tense forms) is longer than the clitics of *would* and *had*. Such a study is already under way in our laboratory, with promising initial results (Homann et al. 2014). One could also test whether inflectional D (as in the past tense) is different in duration from derivational D (as in *three-wheeled*, see Bauer et al. 2013: 304, 306, 313).

If there are indeed systematic differences between the different types of S in speech production, one would also like to know whether language users are influenced by these differences in perception. Walsh & Parker (1983) briefly report a follow-up perception experiment, but there is no proper statistical analysis that would support their conclusion that length does not serve as a perceptual cue.

A natural extension of the present study would therefore be the investigation of S with listeners. The difference in absolute duration between the estimated means of the shortest morphemic S (i.e. the *has* clitic) and non-morphemic S in our

corpus data amounts to approximately 38 ms. This difference seems large enough to be potentially perceptible: the perceptual threshold for durational differences in fricatives has been estimated at about 25 to 30 ms (e.g. Klatt & Cooper 1975, Shatzman & McQueen 2006).

To summarize, this paper presents the first larger study that has systematically investigated the relationship between morphemic status and phonetic implementation of homophonous affixes and their non-morphemic counterpart. This was done using natural conversation data. The analysis has yielded important evidence on the question of affix homonymy, revealing that phonologically homophonous bound morphemes can be phonetically distinct, and that morphemic and non-morphemic S may differ, too. This is unpredicted by current linguistic and psycholinguistic theories of the lexicon and grammar. Further studies are certainly called for to replicate the observed effects, and to develop new models of the mental lexicon and of the relationships between morphology, phonology and phonetic implementation.

REFERENCES

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

Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59.4, 390–412.

Baayen, R. Harald & Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3.2, 12–28.

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1995. *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium.

Baker, Rachel, Rachel Smith & Sarah Hawkins. 2007. Phonetic differences between mis- and dis- in English prefixed and pseudo-prefixed words. *The 16th International Congress of Phonetic Sciences*.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. http://CRAN.R-project.org/package=lme4.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60.1, 92–111.

Berkovits, R. 1993. Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech* 36.1, 89–98.

Boersma, Paul & David J. M. Weenink. 2013. Praat: Doing phonetics by computer (Version 4.3.14) (2005). http://www.praat.org.

Box, George E. P. & David R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society, Series B* 26.2, 211–252.

Bybee, Joan L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan & Paul J. Hopper (eds.). 2011. *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.

Byrd, Dani, Jelena Krivokapic & Sunqbok Lee. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustic Society of America* 120.3, 1589–1599.

Cho, Taehong. 2001. Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica* 58, 129–162.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Coetzee, Andries W. & Joe Pater. 2011. The place of variation in phonological theory. *The handbook of phonological theory*, 2nd edn. 401–434.

Cooper, William E. 1976. Syntactic control of timing in speech production: A study of complement clauses. *Journal of Phonetics* 4.2, 151–171.

Cooper, William E., Jeanne M. Paccia & Steven G. Lapointe. 1978. Hierarchical coding in speech timing. *Cognitive Psychology* 10.2, 154–177.

Crawley, Michael J. 2002. *Statistical computing: An introduction to data analysis using S-Plus*. Chichester: John Wiley.

Cribari-Neto, Francisco & Achim Zeileis. 2010. Beta Regression in R. *Journal of Statistical Software* 34.2, 1–24.

Davies, Mark. 2008. The corpus of contemporary American English: 400+ million words, 1990– present. http://www.americancorpus.org/.

Drager, Katie. 2011. Sociophonetic variation and the lemma. *Journal of Phonetics* 39.4, 694–707.

Ernestus, Mirjam & Natasha Warner. 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics* 39, 253–260.

Ferrari, Silvia & Francisco Cribari-Neto. 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31.7, 799–815.

Foulkes, Paul, Gerard Docherty & Dominic Watt. 2005. Phonological variation in child-directed speech. *Language* 81.1, 177–206.

Fowler, Carol A. 1988. Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech* 31.4, 307–317.

Fowler, Carol A. & Jonathan Housum. 1987. Talkers' signalling of new and old words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26.5, 489–504.

Fromont, Robert & Jen Hay. 2008. Onze miner: The development of a browser-based research tool. *Corpora* 3.2, 173–193.

Gahl, Susanne. 2008. *Time* and *thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84.3, 474–496.

Gahl, Susanne & Alan C. L. Yu. 2006. Special issue on exemplar-based models in linguistics. *The Linguistic Review* 23.3, 213–216.

Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66.4, 789–806.

Goad, Heather. 1998. Plurals in SLI: Prosodic deficit or morphological deficit? *Language Acquisition* 7.2-4, 247–284.

Goad, Heather, Lydia White & Jeffrey Steele. 2003. Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations? *The Canadian Journal of Linguistics/La revue canadienne de linguistique* 48.2, 243–263.

Goldinger, Stephen D. 1998. Echoes of echos? An episodic theory of lexical access. *Psychological Review* 105.2, 251–279.

Hanique, Iris & Mirjam Ernestus. 2012. The role of morphology in acoustic reduction. *Lingue e Linguaggio* 11.2, 147–164.

Hanique, Iris, Mirjam Ernestus & Barbara Schuppler. 2013. Informal speech processes can be categorical in nature, even if they affect many different words. *The Journal of the Acoustical Society of America* 133.3, 1644–1655.

Herberich, Esther, Johannes Sikorski, Torsten Hothorn & Fabio Rapallo. 2010. A robust procedure for comparing multiple means under heteroscedasticity in unbalanced designs. *PLoS ONE* 5.3, e9788.

Holling, Heinz. 1983. Suppressor structures in the general linear model. *Educational and Psychological Measurement* 43.1, 1–9.

Homann, Julia, Ingo Plag & Gero Kunter. 2014. Against homophony: The acoustic properties of English S and D morphemes. *3rd International Society of the Linguistics of English Conference*, August 24–27, Zurich.

Hothorn, Torsten, Frank Bretz & Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50.3, 346–363.

Johnson, Keith. 2004. Massive reduction in conversational American English. *Spontaneous speech: Data and analysis. The 1st Session of the 10th International Symposium*, 29–54. Tokyo and Japan.

Jurafsky, Daniel, Allan Bell, Michelle Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee & Hopper (eds.), 229–254.

Jurafsky, Daniel, Alan Bell & Cynthia Girand. 2002. The role of the lemma in form variation. *Laboratory Phonology* 7, 3–34.

Keating, Patricia A. 2006. Phonetic encoding of prosodic structure. In Jonathan Harrington & Marija Tabain (eds.), *Speech production: Models, phonetic processes, and techniques*, 167–186. New York: Psychology Press.

Keating, Patricia & Stephanie Shattuck-Hufnagel. 2002. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics* 112–156.

214

Kemps, Rachel, Lee H. Wurm, Mirjam Ernestus, Robert Schreuder & Harald Baayen. 2005b. Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes* 20.1-2, 43–73.

Kemps, Rachèl, Mirjam Ernestus, Robert Schreuder & R. Harald Baayen. 2005a. Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory & Cognition* 33.3, 430–446.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In In-Seok Yang (ed.), *Linguistics in the morning calm: Selected papers from SICOL*, 3–91. Seoul: Hanshin.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59.5, 1208–1221.

Klatt, Dennis H. & William E. Cooper. 1975. Perception of segment duration in sentence contexts. In Antonie Cohen & Sibout G. Nooteboom (eds.), *Structure and process in speech perception*, 69–89. Berlin: Springer.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2014. lmerTest. http://cran.r-project.org/web/packages/lmerTest/index.html.

Lavoie, Lisa. 2002. Some influences on the realization of for and four in American English. *Journal of the International Phonetic Association* 32.02, 175–202.

Levelt, William J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Levelt, William J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.01, 1–38.

Levelt, William J. M. & Linda R. Wheeldon. 1994. Do speakers have access to a mental syllabary? *Cognition* 50.1, 239–269.

Li, Hsieh, Laurence Leonard & Lori Swanson. 1999. Some differences between English plural noun inflections and third singular verb inflections in the input: The contribution of frequency, sentence position and duration. *Journal of Child Language* 26.03, 531–543.

Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35.11, 1773–1781.

Losiewicz, Beth L. 1992. *The effect of frequency on linguistic morphology*. Ph.D. dissertation, University of Texas, Austin.

Marian, Viorica, James Bartolotti, Sarah Chabal, Anthony Shook & Stephanie Ann White. 2012. CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE* 7.8, e43230.

Myers-Scotton, Carol & Janice Jake. 2000. Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics* 38.6, 1053–1100.

Nespor, Marina & Irene Vogel. 2007. *Prosodic phonology*. Berlin and New York: Walter de Gruyter.

Nooteboom, Sieb G. 1972. *Production and perception of vowel duration: A study of the durational properties of vowels in Dutch*. Utrecht: University of Utrecht.

Oh, Grace E. & Melissa A. Redford. 2012. The production and phonetic representation of fake geminates in English. *Journal of Phonetics* 40.1, 82–91.

Oller, D. K. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustic Society of America* 54.5, 1235–1247.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee & Hopper (eds.), 137–158.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology VII*, 101–140. Berlin: Mouton de Gruyter.

Pisoni, David B. & Susannah V. Levi. 2009. Some observations on representations and representational specificity in speech perception and spoken word recognition. In Gareth Gaskell (ed.), *Oxford handbook of psycholinguistics*, 3–18. Oxford: Oxford University Press.

Pitt, Mark A., Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume & Eric Fosler-Lussier. 2007. *Buckeye corpus of conversational speech (2nd release)*. Columbus, OH: Department of Psychology, Ohio State University.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Pluymaekers, Mark, Mirjam Ernestus & Harald Baayen. 2005a. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62.2-4, 146–159.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005b. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118.4, 2561–2569.

Pluymaekers, Mark, Mirjam Ernestus, R. Harald Baayen & Geert Booij. 2010. Morphological effects in fine phonetic detail: The case of Dutch -igheid. In Cécile Fougeron (ed.), *Laboratory phonology 10* (Phonology and Phonetics), 511–531. Berlin and New York: Mouton de Gruyter.

R Development coreteam. 2011. R: A language and environment for statistical computing. http://cran.r-project.org/.

Scheer, Tobias. 2010. *A guide to morphosyntax–phonology interface theories: How extra-phonological information is treated in phonology since Trubetzkoy's Grenzsignale*. Berlin, Boston: Walter de Gruyter.

Schwarzlose, Rebecca & Ann R. Bradlow. 2001. What happens to segment durations at the end of the word? *Journal of the Acoustical Society of America* 109.5, 2292.

Shatzman, Keren B. & James M. McQueen. 2006. Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics* 68.1, 1–16.

Song, Jae Yung, Katherine Demuth, Karen Evans & Stefanie Shattuck-Hufnagel. 2013. Durational cues to fricative codas in 2-year-olds' American English: Voicing and morphemic factors. *The Journal of the Acoustical Society of America* 133.5, 2931–2946.

Sugahara, Mariko & Alice Turk. 2009. Durational correlates of English sublexical constituent structure. *Phonology* 26.03, 477–524.

Sugahara, Mariko & Alice E. Turk. 2004. Phonetic reflexes of morphological boundaries at a normal speech rate. In Bernard Bel & Isabelle Marlien (eds.), *Speech Prosody 2004*, 353–356.

Torreira, Francisco & Mirjam Ernestus. 2009. Probabilistic effects on French [t] duration. *INTERSPEECH*, 448–451.

Turk, Alice E. & Laurence White. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* 27.2, 171–206.

Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35.4, 445–472.

Umeda, Noriko. 1977. Consonant duration in American English. *Journal of the Acoustical Society of America* 61.3, 846–858.

Venables, William N. & Brian D. Ripley. 2002. *Modern applied statistics with S-Plus*, 4th edn. New York: Springer.

Walsh, T. & F. Parker. 1983. The duration of morphemic and non-morphemic /s/ in English. *Journal of Phonetics* 11.2, 201–206.

Wurm, Lee H. & Sebastiano A. Fisicaro. 2014. What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language* 72, 37–48.

Yung Song, Jae, Katherine Demuth, Karen Evans & Stefanie Shattuck-Hufnagel. 2013. Durational cues to fricative codas in 2-year-olds' American English: Voicing and morphemic factors. *The Journal of the Acoustical Society of America* 133.5, 2931–2946.

Zeileis, Achim. 2004. Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software* 11.10, 1–17.

Zeileis, Achim. 2006. Object-oriented computation of sandwich estimators. *Journal of Statistical Software* 16.9, 1–16.

Author's address: *(Plag)*
Institut für Anglistik und Amerikanistik,
English Language and Linguistics,
Heinrich-Heine-Universität Düsseldorf,
D-40204 Düsseldorf, Germany
ingo.plag@uni-duesseldorf.de

(Homann)
Institut für Anglistik und Amerikanistik,
English Language and Linguistics,
Heinrich-Heine-Universität Düsseldorf,
D-40204 Düsseldorf, Germany
julia.homann@uni-duesseldorf.de

(Kunter)
Institut für Anglistik und Amerikanistik,
English Language and Linguistics,
Heinrich-Heine-Universität Düsseldorf,
D-40204 Düsseldorf, Germany
gero.kunter@uni-duesseldorf.de