

Variable stem-final fricative voicing in American English plurals:

Different pa[ð ~ θ]s of change

Laurel MacKenzie

New York University

Department of Linguistics

10 Washington Place

New York, NY 10003

212-998-7958

laurel.mackenzie@nyu.edu

Short title: Stem-final fricative voicing in English plurals

Variable stem-final fricative voicing in American English plurals:

Different pa[ð ~ θ]s of change

Abstract This paper investigates analogical leveling in a small set of English nouns which have irregular plural forms. In these nouns, all of which end in a voiceless fricative, the fricative standardly voices in the plural (e.g. wolf – wol[v]es, path – pa[ð]s, house – hou[z]es). Using audio data from three large spoken corpora of American English, I demonstrate that this stem-final fricative voicing is variable and conditioned by a number of factors, most notably the identity of the stem-final fricative — with /f/-final lexemes (e.g. wolf), /θ/-final lexemes (e.g. path), and the /s/-final lexeme house all patterning differently in apparent time — and the frequency of a lexeme in its plural form. I argue that the way these two factors affect the variation is reminiscent of the patterns seen in children’s first language acquisition errors, providing a potential source for the variation and underscoring the importance of considering morphophonological factors when accounting for patterns of change.

Keywords analogical leveling, morphological variation and change, history of English, frequency

Acknowledgements Thanks to Lucy Giannasi and Amy Hughes for data collection, supported by the University of Manchester’s Q-Step Centre. Thanks also to Charles Yang, Isaac Bleaman, members of NYU’s Sociolinguistics Lab, audiences at the 2nd Edinburgh Symposium on Historical Phonology and NWAV 45, and two anonymous reviewers for many helpful comments.

1 Introduction

Formation of the English plural typically involves affixing a /-z/ suffix to a noun. For most nouns, this /-z/ suffix undergoes predictable changes in phonological realization depending on the phonology of the segment which precedes it: after a sibilant fricative, the suffix takes a syllabic allomorph (e.g. ash [æʃ] – ashes [æʃ-əz]); after a voiced non-sibilant segment, it takes a voiced, non-syllabic allomorph (e.g. rig [ɹɪg] – rigs [ɹɪg-z]); and after a voiceless non-sibilant segment, it takes a voiceless, non-syllabic allomorph (e.g. cat [kæt] – cats [kæt-s]). In the majority of English plural nouns, then, the phonological shape of the plural suffix is determined by the phonology of the preceding segment, and the noun stem itself does not change.

However, there is a small set of nouns, all of which end in a voiceless anterior fricative, to which this /-z/ suffix attaches in an irregular way. Rather than the suffix assimilating to the stem-final segment, the stem-final segment assimilates to the suffix. These nouns show a pattern of **rightward voicing assimilation**, which I will henceforth call “stem-final fricative voicing.” Three examples are given in (1), showing the three fricatives that undergo this alternation, /s f θ/.

(1) Rightward voicing assimilation of the stem-final fricative in the plurals of certain English nouns (“stem-final fricative voicing”)

- a. sg. house [haʊs] – pl. houses [haʊz-əz]
- b. sg. leaf [lif] – pl. leaves [liv-z]
- c. sg. path [pæθ] – pl. paths [pæð-z]

This voicing alternation affects only a subset of English nouns ending in /s f θ/: nouns such as kiss, cuff, and death do not undergo it (*[kɪz-əz], *[kʌv-z], *[dεð-z]).

Anecdotal reports indicate that stem-final fricative voicing occurs variably, and may be being lost over time, yielding, e.g., [haʊs-əz], [lif-s], and [pæθ-s] (Becker, Nevins, and Levine 2012; Ringe and Eska 2013). This would constitute a familiar process of regularization, or analogical

leveling. However, these anecdotal reports have not yet been tested on natural speech data. This paper aims to fill this gap by investigating the possibility of variation and change in the voicing of stem-final voiceless fricatives among the small set of English nouns that traditionally undergo this alternation. Specifically, I use large audio corpora to examine the phonological shape of these plural nouns in natural speech, aiming to answer the following questions:

1. Can we confirm previous anecdotal reports that stem-final fricative voicing is variable?
2. If yes, what social and linguistic factors condition the variation?
3. And is this variation part of ongoing change?

After summarizing the historical developments that led to present-day stem-final fricative voicing (Section 2) and outlining the methodology of the present corpus study (Section 3), I demonstrate that this voicing is indeed variable (Section 4). Moreover, it shows phonological class-specific behavior by which /f/-final and /θ/-final nouns show stable variation in voicing in apparent time (though at different overall rates), while houses, the only plural of an /s/-final noun considered here, is rapidly losing voicing in apparent time. Additional factors that affect stem-final fricative voicing, in some cases interacting with the identity of the stem-final fricative itself, are the frequency of a noun in its plural form, the morphological complexity of a plural form, proper nounhood, voicing of a following segment, speaker level of education, and speaking style.

In Section 5, I argue that the fricative-specific and frequency-sensitive patterning of variable stem-final voicing is reminiscent of patterns seen in young children's production of irregular forms elsewhere in the English lexicon. Accordingly, I close by discussing the possible role of children's regularization errors in shaping the present-day patterns of stem-final fricative voicing in adult language. The broader take-home message is that careful consideration of the morpho-phonological conditions on variation can enrich our understanding of a variable phenomenon.

2 Background

2.1 History of stem-final fricative voicing

The voicing alternation presented in (1) has its source in two characteristics of Old and Early Middle English, both of which have been lost in the present day (Lass 2000; Ringe and Eska 2013). First, the plural suffix was uniformly the syllabic morpheme /-əs/, with none of the allomorphic alternations attested today. Second, there was a regular, allophonic process of voicing of anterior fricatives (/f s θ/) intervocalically. As a result, nouns such as [wɔlf] ‘wolf’, [paθ] ‘path’, and [hu:s] ‘house’ formed the regular plurals [wɔlv-əs] ‘wolves’, [pað-əs] ‘paths’, and [hu:z-əs] ‘houses’. This voicing alternation was also observable elsewhere in the language, for instance in the formation of denominal verbs such as [bað-ən] ‘to bathe’ and [hu:z-ən] ‘to house’ from the nouns [baθ] ‘bath’ and [hu:s] ‘house’.

As Ringe and Eska (2013:142–144) explain, a series of contact-induced and language-internal changes between the twelfth and fifteenth centuries created a number of lexical exceptions to the productive rule of intervocalic anterior fricative voicing. These exceptions, which contained voiceless anterior fricatives in intervocalic position, likely led to the loss of intervocalic voicing as a productive rule. Nevertheless, a number of forms which were historically subject to the rule ([wɔlv-əs], [pað-əs], [hu:z-əs], and so forth) preserved their voicing, presumably newly marked as exceptions themselves.¹ This means that, around the fifteenth century, stem-final fricative voicing in a plural form like wolves changed from having its source in a regular and productive rule of intervocalic voicing, to having its source in a lexical listing pairing the singular wolf with the voiced plural wolves, in a similar way to how the noun child in Present-Day English is lexically listed as having the exceptional plural children.

Presumably, speakers today still have an irregular rule voicing the stem-final fricative in the plural of this small set of exceptional nouns, and a much more regular and productive rule that

affects the plural /-z/ suffix in other plural forms. The question addressed in this paper is the extent to which the items in the irregular voicing set still undergo it today: in other words, whether analogical leveling with voiceless models has regularized them.

2.2 Present-day status of stem-final fricative voicing

Lass (2000), Ringe and Eska (2013), and Jespersen (1942) all mention that at least some items in the voicing set voice only variably. Lass (2000:145) further describes stem-final fricative voicing as “recessive,” implying that it may be being lost diachronically. However, no researcher has yet carried out an empirical study of English speakers’ production of stem-final fricative voicing among the nouns that have historically undergone it.

The closest we get to a production study of present-day stem-final fricative voicing is a judgment study by Becker et al. (2012). In this study, English-speaking subjects heard two plural forms for a number of /f/-final and /θ/-final nouns and were asked which they preferred: one with stem-final voicing (e.g. [pæðz] ‘paths’) or one without (e.g. [pæθs]). Stimuli included nouns that have been historically subject to stem-final fricative voicing in the plural (e.g. calf, bath) as well as nouns that have not (e.g. fife, cough), with this latter category including a number of polysyllabic items (e.g. giraffe, zenith). Becker et al. find significant effects of syllable count, stress pattern, and preceding vowel on speakers’ preferences for stem-final voicing: speakers prefer voicing in the plurals of monosyllabic items, items containing long vowels, and, when an item is polysyllabic, in iambs rather than trochees. However, Becker et al. do not separate out those items for which stem-final fricative voicing is attested historically (e.g. calf, bath) from those for which it is not (e.g. fife, cough, giraffe, zenith), and it is very possible that at least the monosyllabic effect is due to this confound: the stems that traditionally or standardly undergo stem-final fricative voicing, according to Jespersen and Ringe and Eska, are all monosyllables.

Becker et al. also find no effect in their study of a number of factors that turn out to matter in the present study, namely the identity of the stem-final fricative, an item's frequency, its morphological complexity, and a speaker's age (Section 4). This is perhaps unsurprising, since, to reiterate, they considered a larger set of nouns than just those that have been historically subject to stem-final fricative voicing.

As it stands, then, no study has examined the production of stem-final fricative voicing in natural speech, and the one study that has collected empirical data on this phenomenon has not looked at it from the perspective of language change. In order to answer the research questions listed in Section 1, this paper presents data from three large audio corpora of present-day American English speakers' pronunciations of the words that are traditionally subject to stem-final fricative voicing.

3 Methodology of the corpus study

3.1 Defining the variable

The set of nouns which undergo stem-final fricative voicing in the plural has changed over time. Some nouns have definitively left the class of those that undergo the alternation: for instance, Jespersen (1942) cites the archaic form cleves as a historical plural of cliff (with stem-final fricative voicing plus a vowel change, neither of which are attested in this form today). Others are noted by Jespersen as having joined the voicing set, including moth and scarf. Moreover, Becker et al. (2012) cite anecdotal reports of stem-final fricative voicing in the plurals of giraffe and photograph, and Hayes (2011) of stem-final fricative voicing in the plurals of gulf, chief, and epitaph. Thus it is very difficult to know a priori which nouns undergo fricative voicing in the plural without carrying out a corpus study of the pronunciation of the plurals of every /f/-final noun in English, if not also the /s/-final and /θ/-final ones.

This was untenable for the current paper. Instead, the decision was made to examine only the set of nouns identified by Jespersen (1942) as either variably or categorically undergoing stem-final fricative voicing in the plural. This restricts the present investigation to the **maintenance or loss** of fricative voicing from those items that were traditionally subject to it, rather than the present-day extension of fricative voicing to items that were not traditionally subject to it. The latter topic is nonetheless a question of considerable interest, as there are several reasons to suspect that it may be occurring. These include the anecdotal reports of expansion by Becker et al. (2012) and Hayes (2011) cited earlier, the historical precedent for such an expansion by forms like scarves, and a brief note by Jespersen (1942:264) that certain British English dialects may have widespread fricative voicing in the plurals of nouns that were not historically subject to the anterior fricative voicing rule, such as faces, places, and prices. However, this is a different question from that of maintenance or loss of fricative voicing among those nouns that traditionally underwent it, so I leave it for future research.

The 33 nouns listed in (2) were selected for study, due to Jespersen's (1942:258–264) identification of them as either categorically or variably undergoing stem-final fricative voicing in the plural.

- (2) Nouns identified by Jespersen (1942:258–264) as undergoing stem-final fricative voicing in the plural
- a. /f/-final: beef, calf, dwarf, elf, half, hoof, knife, leaf, life, loaf, oaf, roof, scarf, self, sheaf, shelf, staff (in the meaning 'stick'), thief, turf, wharf, wife, wolf²
 - b. /s/-final: house
 - c. /θ/-final: bath, cloth, lath, moth, mouth, oath, path, sheath, truth, wreath, youth

Throughout this paper, I refer to these nouns as “lexemes” and the three phonological group-

ings they form (/f/-final, /s/-final, /θ/-final) as “phonological subsets” or “classes.” In graphs and throughout the text, I present the lexemes given in (2) in their singular form. This is because standard spelling conventions require that plural forms of the /f/-final lexemes be spelled as if stem-final fricative voicing has applied, and this may be confusing in a paper that demonstrates that this voicing often in fact has not applied. Despite this choice, what is at issue in this paper is speakers’ production of the plural forms of these lexemes.

3.2 Data collection

Three corpora were selected for study:

- Switchboard (Godfrey and Holliman 1997): A corpus of short telephone conversations between strangers on assigned topics, collected between 1991 and 1992. Switchboard consists of 2,400 such conversations collected from 542 different speakers, amounting to over 250 hours of speech/ca. 3 million transcribed words.
- Fisher (Cieri 2004): Like Switchboard, a corpus of short telephone conversations between strangers on assigned topics, collected between 2002 and 2003. Fisher consists of over 11,000 such conversations, amounting to ca. 2,000 hours of speech.
- The Philadelphia Neighborhood Corpus (“PNC,” Labov and Rosenfelder 2011): A corpus comprising 40 years’ worth of sociolinguistic interviews with 318 different native Philadelphians, carried out by trained student fieldworkers at the University of Pennsylvania between 1973 and 2012. The PNC contains approximately 150 hours of speech/ca. 1.6 million transcribed words.

These corpora were chosen because they are large, transcribed, time-aligned, and all reflect conversational speech.

Transcripts of all three corpora were searched for plural forms of the 33 lexemes listed in (2).

Where the presence or absence of stem-final fricative voicing is encoded in a plural form's spelling (i.e. for /f/-final lexemes, which vary in the orthography of their plurals between final <-ves> and final <-fs>), both possible spellings were searched for. All searches included a wildcard character preceding the lexeme, in order to bring up single-word compound forms (e.g. were-wolves, warehouses, tablecloths) where they occurred. A Praat script (Boersma and Weenink 2015) was then used to match each hit in a transcript to its corresponding occurrence in the audio for subsequent auditory coding (on which see Section 3.3.1).

Of the 33 lexemes listed in (2), no hits were found for the plurals of oaf, sheaf, staff, turf, wharf, lath, or sheath, so no data on these lexemes will be reported. One lexeme was additionally omitted from study: self. The plural of this item turned out to occur incredibly infrequently except as part of the anaphoric pronouns themselves, ourselves, and yourselves. As part of those pronouns, the lexeme was overwhelmingly frequent: 3,854 instances of -selves pronouns across the three corpora compared to 41 instances of the bare plural selves. Such an imbalanced ratio of bimorphemic forms to monomorphemic forms is not found for any other lexeme in (2). The closest is cloth, which occurs six times in the data in a compound form (e.g. tablecloths) compared to one time in its bare form cloths, but a 6:1 ratio is a far cry from the 94:1 ratio displayed by selves. Self is thus an outlier among the lexemes in (2) for its overwhelming frequency in a compound plural form.

Additionally, a second characteristic makes self an outlier: the bimorphemic forms in which it overwhelmingly occurs are all function words (specifically, pronouns). No other lexeme in (2) is or forms part of a function word. For these reasons — its near-exclusive function word status, its near-exclusive bimorphemicity, not to mention its extreme frequency, dwarfing the second-most frequent plural form in the data, houses, by four times — self was omitted from study, as

any one of these factors could potentially contribute to its being phonologically different from the other lexemes in (2). I leave a detailed study of stem-final fricative voicing of self and its derivatives to future work, though I return to some pilot data on these forms in Section 5.

Before auditory coding, search results were culled to omit items that were homographs with the plural forms under study but represented a different lexeme (e.g. the verb lives, the verb leaves). Similarly excluded from study were items where a string at the end of a plural form was a homograph with one of the plural nouns under study but bore no etymological relation to it (e.g. the moths in mammoths; the paths in psychopaths), again to ensure that the items under study were only those that were historically subject to voicing.

In all, 2,361 tokens of the 33 lexemes given in (2) are analyzed here. Table 1 breaks this down by stem-final fricative.

	f	s	θ
N	1,154	1,048	159

Table 1: Total number of tokens obtained for each stem-final fricative.

3.3 Coding

3.3.1 Dependent variable

Two undergraduate research assistants at the University of Manchester listened to every token in the data individually³, coding each one as voiced, voiceless, or reject (due to a token being unclear, ambiguous, or mistranscribed). They then met to compare their coding. Rates of agreement for their judgments of voiced vs. voiceless are 88% ($\kappa = 0.76$) for each of Fisher and the PNC, and 84% ($\kappa = 0.68$) for Switchboard; these results reflect good reliability, according to Clopper (2011). The two coders then re-listened to each token on which they disagreed, and either arrived at an agreement, or rejected the token from study. As such, the results presented here constitute only those instances in which both coders agreed on the voiced vs. voiceless judgment.⁴

It is well documented that so-called voiced consonants in English, especially word-final ones, may not be fully voiced articulatorily (see Davidson 2016 for a recent review), and that perception of the so-called voiced/voiceless contrast in final consonants may be better attributed to the differing durations of pre-voiced vs. pre-voiceless vowels than to the presence or absence of consonant voicing (Raphael 1972). This means it is debatable whether the coders were actually coding vocal fold vibration, or rather some other cue. Still, the high rate of agreement between them indicates that, whatever they were hearing, they were hearing the same thing. I continue to refer to the phenomenon under study here as “voicing,” with “voiced” and “voiceless” forms, while acknowledging that vocal fold vibration may not be the operative acoustic measure.

3.3.2 Independent variables

All tokens were coded for the following linguistic and social predictors of stem-final fricative voicing.

- **Identity of stem-final fricative.** Irregular forms have been found to show class behavior, or subregularities, in first language acquisition (Yang 2002) and language processing (Regel et al. 2015). Accordingly, we can ask whether the set of lexemes that voice in the plural divides into phonological subsets: namely, whether /f/-final lexemes behave differently from /θ/-final lexemes and from the /s/-final lexeme house. Items were accordingly coded for the final fricative of their stem: /f/ (as in leaf, wolf), /s/ (exclusively in house), or /θ/ (as in bath, wreath).
- **Stem identity.** Each token was coded for the lexical identity of its stem, to be incorporated as a random effect in statistical modeling (see Section 3.4). For instance, houses, greenhouses, and frat houses were all coded as having the stem house.
- **Morphological complexity.** Compound words typically take regular morphology even when

their monomorphemic stems do not: hence, according to Pinker (1994), the past tense of grandstand is grandstanded, not *grandstood, and the plural of lowlife is lowlives, not *lowlives. Accordingly, we can hypothesize that stems that form the second part of a lexicalized compound will be less likely to voice than those that do not.

Given this, each plural instance of the lexemes in (2) was coded for whether it appeared as the second member of a lexicalized compound (e.g. housewives, birdhouses, tablecloths).

This was assessed as follows: each time a plural form appeared as the second element of a noun–noun or adjective–noun pair, the pair (in its singular form) was looked up in the Oxford English Dictionary. If it appeared as an entry or sub-entry, it was coded as a lexicalized compound, regardless of the specifics of its orthography (hence two-word steak houses, hyphenated half-truths, and single-word bookshelves were all coded as compounds, because they all appear as (sub-)entries in the OED). In rare cases, a pair did not appear in the OED, but the speaker had uttered the pair with compound stress (Chomsky and Halle 1968), i.e. more stress on the first element than the second. (Compare the stress pattern of the lexicalized compound gréenhouses ‘glass buildings where plants are grown’ to that of the non-compound grèen hóuses ‘houses that are green.’) Pairs uttered with compound stress were also coded as compounds. Examples of items that were pronounced with compound stress but are not listed in the OED as lexicalized compounds are drop cloths and frat houses.

For all three phonological subsets (/f/-final items, /s/-final items, and /θ/-final items), it is more common in this data that a lexeme surfaces in its bare form than as part of a lexicalized compound (bare-to-compound ratio for /f/-final lexemes: 14:1, /s/-final lexemes: 6:1, /θ/-final lexemes: 4:1).

- **Proper noun.** The Toronto hockey team is famously the Maple Leafs, not the Maple Leaves

(Pinker 1994). Like compounds, proper nouns are often exempt from irregular morphology. As such, we can hypothesize that proper nouns will be less likely to show stem-final fricative voicing than common ones, so each token was coded for whether it was part of a proper noun (e.g. Snow White and the Seven Dwarfs, Houses of Parliament) or not.

- **Frequency of lexeme in its plural form.** Frequency plays a well-documented role in diachronic leveling and in synchronic regularization among child language learners, with low-frequency items more likely to regularize than high-frequency ones (Hooper 1976; Phillips 1984; Yang 2002). We can thus hypothesize that, if stem-final fricative voicing turns out to be variable, we will find a positive correlation between frequency and maintenance of voicing.

The frequency of each of the lexemes in (2) in its plural form was obtained from the SUBTLEXus corpus (Brysbaert and New 2009), a corpus of word frequencies based on the subtitles of 8,388 American television programs, amounting to 51 million words. Brysbaert and New find these frequency measures to do a better job of predicting lexical decision times for speakers of American English than any previously-used frequency metrics.

Where a plural form could be spelled in two different ways depending on whether it has undergone stem-final fricative voicing or not (namely, any /f/-final lexeme, e.g. wolves ~ wolfs), frequencies of each of the two possible spellings were obtained and summed together. (Unsurprisingly, given the fact that stem-final fricative voicing is standard in the plurals of most of the lexemes listed in (2a), the <-ves> spelling generally turned out to be much more frequent than the <-fs> spelling; sometimes the <-fs> spelling was not even attested in SUBTLEXus.)

Frequencies were transformed to the Zipf scale presented in van Heuven et al. 2014. This

scale takes the \log_{10} of an item's frequency per billion words. As van Heuven et al. discuss, this provides an intuitive logarithmic scale from 1 (very low frequency) to 7 (very high frequency). To improve regression model convergence, Zipf values were centered around the median for statistical analysis.

It is important to note that for all items in the data, it is the plural frequency of the stem, not necessarily the plural frequency of the word, that is used to test for effects of frequency on variation and change in stem-final fricative voicing. That is, when a plural appeared as part of a lexicalized compound (e.g. steak houses, half-truths, bookshelves), it is the frequency of the stem in its plural form (houses, truths, shelves) that was used in modeling. This is because SUBTLEXus provides no frequency measures for two-word or hyphenated items, only items that are standardly written as a single word without a space. Steak houses and half-truths (and, for that matter, steakhouses and halftruths) do not appear in SUBTLEXus, so their frequencies cannot be obtained, even though the frequency of a conventionally single-word compound like bookshelves can. For this reason, plural stem frequency rather than plural word frequency was used to test the role of frequency in stem-final fricative voicing. Two of the lexemes in (2) turn out to be polysemous: calf, which can refer to either the animal or the muscle, and bath, which can refer to either the method of cleansing or the bathroom in a house (as in “three beds, two baths”). Both meanings of each lexeme appear in the data. SUBTLEXus does not provide separate frequency measures for each meaning of a lexeme, so the frequency of these items was simply coded as the general overall frequency of calves+calfs and baths, respectively.

- **Voiced counterpart.** Paradigm uniformity effects have been found to play a role in synchronic variation: a word's phonetic realization can be influenced by the phonetic realiza-

tion of morphologically related words (see Seyfarth 2016 for a recent review). Accordingly, we can hypothesize that lexemes which have an invariably voiced counterpart elsewhere in the English lexicon (e.g. bath with /θ/ has the counterpart bathe with /ð/) will be more likely to show fricative voicing in the plural than lexemes which do not (e.g. path).

To test this, each token was coded for whether its stem has a voiced counterpart elsewhere in the English lexicon (typically a verb). Of the lexemes listed in (2), the following were identified as having a voiced counterpart: bath (bathe), calf (calve), cloth (clothe), half (halve), house (house (v.)), life (live (v.), live (adj.)), mouth (mouth (v.)), shelf (shelve), thief (thieving, thievery), wreath (wreathe).

- **Voicing of following segment.** Devoicing of English /z/ is well attested in two environments: phrase-finally (e.g. leaves ♪), and before a following voiceless obstruent (e.g. leaves fall) (Smith 1997; Myers 2010). This means that there may be instances in which a speaker intends to utter a form with stem-final fricative voicing, but, due to devoicing of the /z/ plural suffix followed by concomitant devoicing of the stem-final fricative that precedes it (itself due to English phonotactic restrictions on voicing agreement in final obstruent clusters, Hayes 2011), the token comes out sounding as if voicing has not applied. So, we can hypothesize that plural nouns which occur phrase-finally or before a voiceless consonant will be less likely to have been heard as having stem-final fricative voicing than plural nouns which occur before a voiced segment. To this effect, tokens were coded for whether the first segment of the word following the plural noun was voiced or voiceless, or whether there was no following word (due to the speaker pausing or ending their turn). Voiced/voiceless coding was obtained automatically from the CMU Pronouncing Dictionary 0.7 (Weide 2008); where a word was not in the dictionary, voicing was input by hand.

It is worth noting that we would expect cross-word assimilatory devoicing to be at play for /f/-final and /θ/-final items, but not for tokens of houses and its derivatives. In houses, a vowel separates the stem-final fricative from the plural /z/ suffix. Even if the /z/ suffix devoices, we wouldn't expect concomitant devoicing of the stem-final fricative in house, which does not form a final obstruent cluster with the plural suffix.

- **Speaker year of birth.** Speaker year of birth, as recorded in each corpus, was examined as a way of assessing change in apparent time. This measure was centered around the median and rescaled to decades to facilitate regression modeling and interpretation.

In analyzing the data for potential change in progress, the question arose of whether to use speaker year or birth or speaker age at recording (i.e. year of recording – speaker year of birth) as the relevant measure. This question does not arise in the typical apparent-time study, where data comes from recordings taken over a short period of time, and year of birth and age at recording effectively share a one-to-one mapping. But by compiling the forty years of interviews in the PNC with Switchboard and Fisher, we effectively conflate both real and apparent time. A speaker born in 1960 who was interviewed in the PNC in 1975 would have been fifteen at the time of recording, whereas a speaker born in 1960 but interviewed in 2002 for Fisher (or the PNC, for that matter) would have been forty-two. If stem-final fricative voicing is subject to age-grading — change within an individual as they age of a variable which is diachronically stable in the community (Sankoff 2005) — then conflating data from a fifteen-year-old speaker and a forty-two-year-old speaker simply because both were born in 1960 would lead us to lose valuable information.

However, there is no a priori reason to believe that stem-final fricative voicing is subject to age-grading. As Chambers (2009) observes, age-graded changes are rare; moreover, classic

age-graded variables are those where one variant bears a clear and strong social evaluation in the community, such as [ɪŋ] ~ [ɪn] variation and negative concord (Labov 2001). Overt prescriptions concerning stem-final fricative voicing are not attested in the same way that they are for these variables: this variation seems to be generally below the radar, so it is unlikely that it is age-graded, or that we are in danger of obscuring any age-related pattern when we analyze the data by speaker date of birth.⁵

Additionally, data from the PNC only make up 16% of the total data analyzed here: it is a small corpus compared to Switchboard and Fisher. Indeed, 87% of the data in the present study was collected between 1991 and 2004. Despite the inclusion of PNC data, the present study is effectively an apparent-time analysis over a short span of thirteen years. Still, in Section 4.8, where I examine potential change in progress, I confirm that any year-of-birth effect present in the overall data set is also present in the subset of data collected between 1991 and 2004, to get as close to a true apparent-time study as possible without dramatically reducing token counts.

- **Speaker sex.** Each corpus reported each speaker's sex as female or male. One speaker, from Fisher, whose sex was given as NA was omitted.
- **Speaker level of education.** Fisher and the PNC report speaker education level in years of schooling, but Switchboard reports it on a four-point scale (less than high school, less than college (but not less than high school), college, more than college).⁶ This meant that speaker education levels for Fisher and PNC speakers had to be coerced onto Switchboard's scale. To do this, speakers whose years of schooling was given as 8 or less were coded as "less than high school," 12 or less as "less than college," 16 or less as "college," and greater than 16 as "more than college." This was further coerced into "less than college" ("less than high

school” + “less than college”) versus “college or more” (“college” + “more than college”) to deal with small token counts and model convergence problems.⁷

- **Corpus.** Corpus (PNC, Switchboard, or Fisher) was included as a fixed-effect predictor as the different conversational settings of the PNC (sociolinguistic interviews) vs. the other two corpora (phone conversations) may result in different patterns of variation.

3.4 Statistical modeling

Statistical modeling of the data was carried out using mixed-effects logistic regression in R (R Core Team 2016) via the `lme4` package (Bates et al. 2015). Modeling for this study had to be undertaken with care, for a number of reasons. First, token counts are not large, especially for /θ/-final lexemes (Table 1). As a result, particularly complex statistical models (including more than one or two interactions) typically failed to converge. Additionally, a downside of working with corpus data is that not all independent variables will be consistently coded or recorded across the different corpora. This meant that occasionally high-dimension coding had to be reduced to a lower dimension, as in the example of speaker level of education given above, where years of schooling in Fisher and the PNC had to be coerced to Switchboard’s four-point scale.

Modeling proceeded as follows. The data from the three phonological subsets (/f/-final, /s/-final, /θ/-final) was pooled and analyzed in a model that was as maximal as possible while still resulting in successful convergence. This model included all fixed-effect predictors enumerated in Section 3.3.2. It also included a random effect of item. For /f/-final and /θ/-final items, this random effect coded for the stem of the word recorded for that data point (e.g. bath, path, knife, leaf, etc.). (Using the word itself in the case of lexicalized compound items — e.g. flea bath, bike path, exacto knife, bayleaf, etc. — resulted in convergence failure). For /s/-final items, this random effect coded for the actual word recorded, even if it was a lexicalized compound (e.g. fire-

house, opera house, outhouse, etc.). This is because all /s/-final items have the same stem — house — so using stem as the random effect for /s/-final items, as was done for /f/-final and /θ/-final items, would create statistically problematic isomorphy with the fixed effect of stem-final fricative identity.

A random effect of speaker, though common in other corpus-based sociolinguistic studies, was not included, for two reasons. First of all, the data in this study come from 1,506 different speakers; the mean number of tokens per speaker in the data is 1.6 and the median is 1. Random speaker effects are meant to prevent individual speakers who contributed many tokens to a data set from erroneously skewing the results (Johnson 2009), but this single-speaker influence is hardly possible in a data set where no speaker contributed a predominance of tokens. Second, having this additional random effect resulted in an increased likelihood of model convergence failure.

Then, each fixed-effect predictor was assessed for significance in the pooled data as follows. A base model including only the random intercept of item was compared to a model containing that same random intercept and one fixed-effect predictor from the full model in a Chi-square test comparison of models using R's `anova` function. (A predictor which achieved a large effect size and statistically significant p -value in the full model was selected for this step.) This one-predictor model was then compared to a second model containing that same fixed-effect predictor plus one more (again, another predictor that appeared to be accounting for a considerable amount of variation based on the full model) in another Chi-square test ANOVA model comparison. Subsequent comparisons proceeded in this way until all fixed-effect predictors had been tested for the significance of their contribution. The results of this set of comparisons allowed the researcher to create a final model, one which contained only those predictors which made a sig-

nificant ($p < 0.05$) contribution to the model.⁸

Because it has been found that irregular forms often show phonological class behavior (see Section 3.3.2), it was important to assess whether any predictors that affected variation and change in the pooled data operated in the same way on each phonological subset: that is, whether /f/-final lexemes, /s/-final lexemes, and /θ/-final lexemes all patterned in the same way. Including multiple interactions between stem-final fricative identity and the other predictors of interest resulted in pervasive model convergence difficulties. As a result, in addition to the pooled model, separate models were fitted for /f/-final lexemes, /θ/-final lexemes, and houses and its derivatives. These were carried out using the same maximal model → Chi-square test ANOVA comparisons → significant-predictors-only procedure described in the previous paragraph. Each by-fricative model again had a random effect of stem, with the exception of the model of the /s/-final data, which had a random effect of word, due to its comprising only one stem (house).

Throughout Section 4, I discuss the effect of each predictor on the pooled data, then enumerate whether this holds for the three different phonological subsets (/f/-final, /s/-final, and /θ/-final).

4 Results

4.1 General pattern

Rates of stem-final fricative voicing by identity of stem-final fricative are plotted in Figure 1. Quite generally, we can see from this figure that stem-final fricative voicing in the set of lexemes studied here is indeed variable: for no phonological subset does the rate of voicing ever exceed 85%.

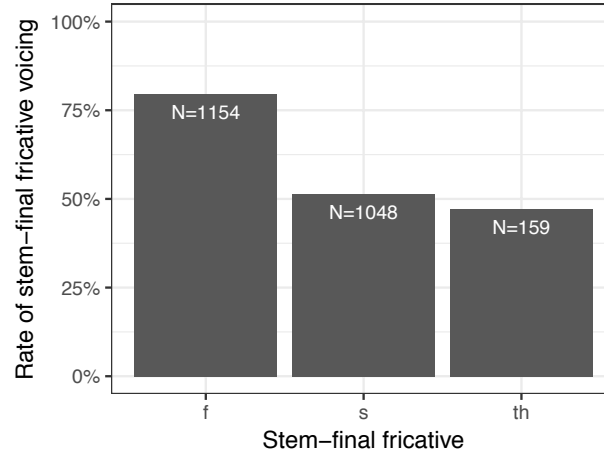


Figure 1: Rate of stem-final fricative voicing by identity of stem-final fricative.

Stem-final fricative voicing rate also clearly varies according to the identity of the stem-final fricative itself: the historically regular voicing rule has fragmented into differently variable rules for the three phonological subsets. The /s/-final and /θ/-final subsets both undergo fricative voicing at a significantly lower rate than the /f/-final subset (/s/-final: $\beta = -0.977$, $p = 0.007$; /θ/-final: $\beta = -0.994$, $p = 0.01$).

4.2 Lexeme-specificity

Figure 2 breaks down the data of Figure 1 by lexeme. There is one dot for each lexeme studied, sized by the number of tokens of that lexeme in the data. The \log_{10} of the token count for each lexeme was taken in order to better visualize vastly different token counts. Lexemes have been sorted from highest to lowest voicing rate. In cases of polysemy, lexemes with different meanings have been plotted separately.

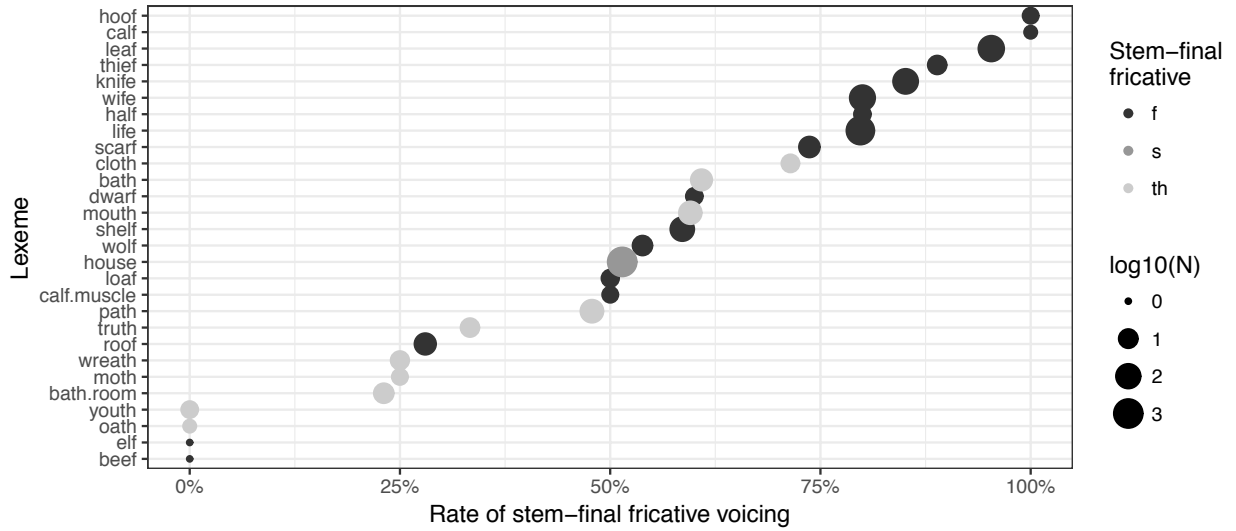


Figure 2: Rate of stem-final fricative voicing by lexeme. Dot size reflects the log10 of the number of tokens of that lexeme in the data. Lexemes have been sorted from highest to lowest voicing rate.

We see from Figure 2 that although /f/-final lexemes (black dots) do undergo stem-final fricative voicing at an overall higher rate than lexemes ending in /θ/ or /s/, not all /f/-final lexemes voice at a uniformly high rate: the lexemes loaf, calf (with its ‘muscle’ meaning), and roof all voice 50% of the time or less.⁹

Figure 2 does show, though, that all /θ/-final lexemes (lightest gray dots) undergo fricative voicing at a fairly low rate. No /θ/-final lexeme shows more than 75% voicing; the highest is cloth, voicing at a rate of 71% (though N = 7).

Figure 2 also shows that the different meanings of polysemous lexemes do not necessarily undergo stem-final fricative voicing at the same rate. For instance, we see considerably different voicing rates for calf with the meaning ‘animal’ vs. calf with the meaning ‘muscle’ (100% vs. 50%, though N = 2 and 4, respectively), and bath with the meaning ‘method of cleansing’ vs. bath as an abbreviated form of bathroom, as when describing the number of rooms in a house (61%

vs. 23%; N = 23 and 13, respectively).

Overall, we see general effects of more voicing across the board of /f/-final lexemes and less voicing of /θ/-final lexemes, but also lexeme-specific behavior within those phonological subsets. This motivates the inclusion of a random effect of lexeme in the logistic regression models used here, as well as testing for the role of other lexeme-specific predictors, namely frequency, as I do in the next section.

4.3 Frequency of lexeme in its plural form

A first indication that the plural frequency of a lexeme may be playing a role in stem-final fricative voicing variation is the finding, presented in the previous section, that /f/-final lexemes voice more than /θ/-final lexemes. The plural forms of the /f/-final lexemes in this data are on average more frequent than the plural forms of the /θ/-final lexemes. (Average Zipf scale frequency, /f/-final: 3.56, /θ/-final: 3.06.) Indeed, as Figure 3 demonstrates, the lexemes in this data with the highest plural form frequency all end in /f/ (apart from the lone /s/-final lexeme house).

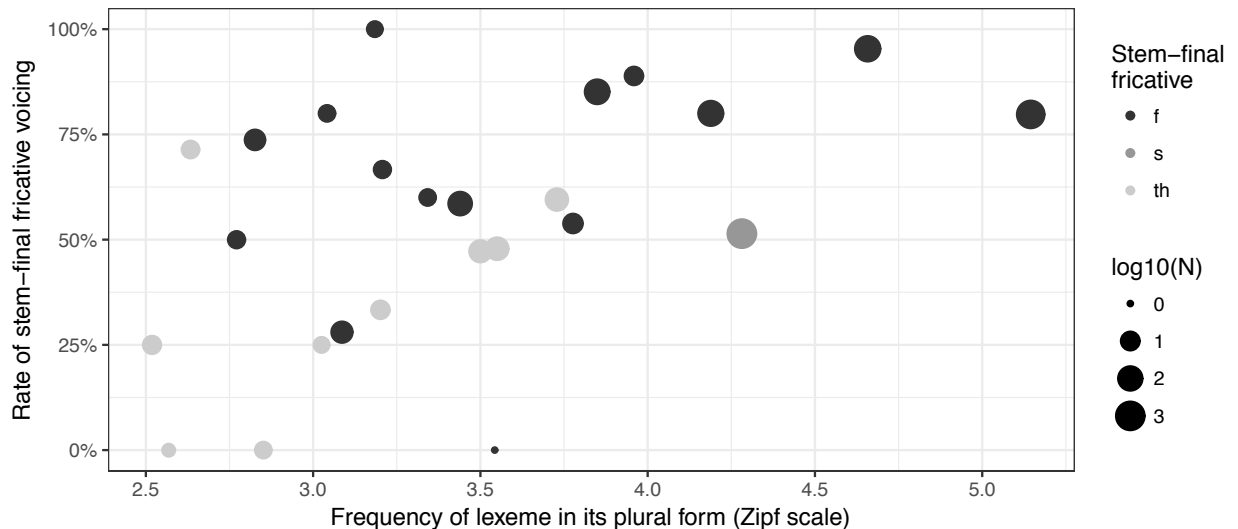


Figure 3: Rate of stem-final fricative voicing by frequency of plural form. One dot per lexeme.

Dot size reflects the log₁₀ of the number of tokens of that lexeme in the data.

Figure 3 plots the rate of stem-final fricative voicing for each lexeme by the frequency of that lexeme, in its plural form, on the Zipf scale.¹⁰ This allows us to visualize the role of plural form frequency within the two categories of /f/-final and /θ/-final lexemes.

Plural form frequency has a significant positive effect on stem-final fricative voicing in the data overall ($\beta = 0.824$, $p = 0.001$), as well as when we subset out the data by stem-final fricative (/f/-final: $\beta = 0.939$, $p = 0.015$; /θ/-final: $\beta = 1.066$, $p = 0.032$). In all cases where frequency achieves significance, its coefficient is positive; that is, the more frequent the plural form of a lexeme is, the more its final fricative voices in the plural.

Overall, then, we see two effects of frequency in the data: (1) on a macro level, /f/-final lexemes, whose plurals are on average more frequent than those of /θ/-final lexemes, voice more; (2) within the subsets of /f/-final and /θ/-final lexemes, lexemes whose plurals are more frequent are more likely to voice.

However, it is important to note that the frequency–voicing correlation is not perfect. Houses has a Zipf scale frequency of 4.28 — higher than the plurals of many /f/-final lexemes — yet it does not display a concomitantly high rate of voicing. I return to this point in Section 5.

4.4 Morphological complexity

Figure 4 plots the rate of fricative voicing by the identity of a lexeme's final fricative and whether or not it appeared as part of a morphologically complex form, in order to assess whether stem-final fricative voicing is indeed less likely in compounds (Pinker 1994).

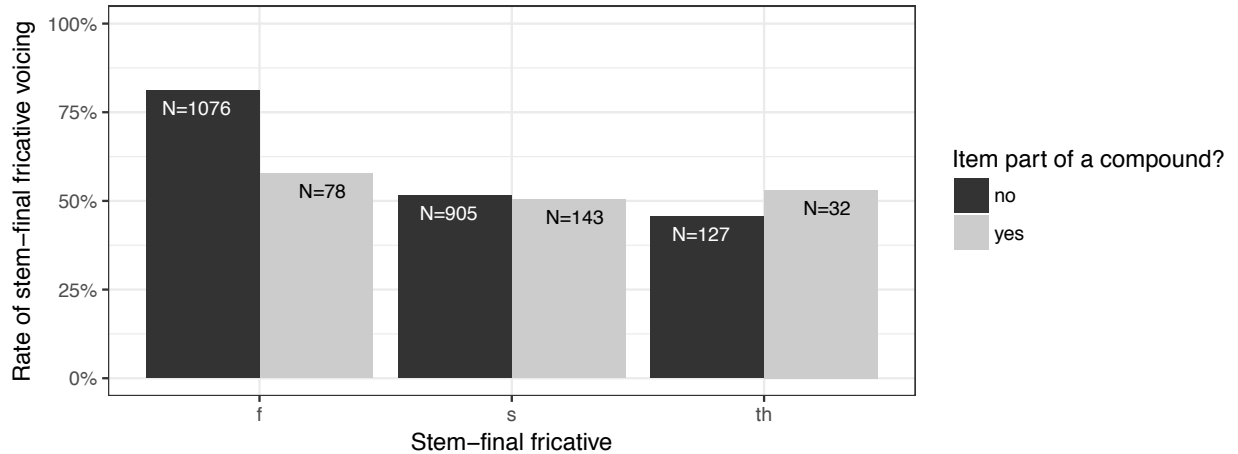


Figure 4: Rate of stem-final fricative voicing by identity of stem-final fricative and item's compound status.

Morphological complexity is a significant negative predictor of stem-final fricative voicing in the pooled data ($\beta = -0.682$, $p = 0.003$), but this effect turns out to be present in the /f/-final subset only ($\beta = -1.146$, $p = 6.1 \times 10^{-5}$); it does not reach significance for the /s/-final or /θ/-final subsets. Only in the plurals of /f/-final lexemes, then, do we find morphological complexity disfavoring voicing.

Why should this be fricative-specific? Pinker's (1994) proposal for why a compound like lowlife doesn't get irregular plural morphology (**lowlives*) is that a lowlife is a type of person, not a type of life; for this reason, the irregular morphology of the component life doesn't percolate up to the compound. Are there more idiomatic /f/-final compounds than /s/-final or /θ/-final compounds? In fact, this appears to be the case. The disfavoring effect of morphological complexity in the /f/-final subset is driven entirely by four items: lowlife, bookshelf, and two compounds which are also proper nouns: Timberwolves (the Minnesota basketball team) and Maple Leafs (the Toronto hockey team). Omitting these four items from the /f/-final data eliminates the significant effect of morphological complexity, even though a large number of compound forms

remain (e.g. exacto knife, pocket knife, personal life, housewife, werewolf, bay leaf). The role of proper nounhood in blocking voicing will be discussed in the next section; as for lowlife and bookshelf, it appears that speakers must be failing to decompose them, as Pinker proposes. (A bookshelf is perhaps seen as a piece of furniture rather than a type of shelf.) So, morphological complexity does play a role in blocking stem-final fricative voicing, but only in those cases in which a compound is opaque. The large number of transparent /s/-final compounds (e.g. birdhouse, courthouse, dollhouse, firehouse, steakhouse, townhouse, warehouse) and /θ/-final compounds (e.g. tablecloth, bike path, career path, half-truth) prevent the emergence of a similar disfavoring effect of morphological complexity in those subsets.

4.5 Proper noun status

Pinker (1994) proposes that speakers will be unlikely to apply irregular morphology to a proper noun because what is pluralized is a name, not a noun, and a plural must be based on a noun, rather than a name, in order to receive that noun's irregular morphology. Stem-final fricative voicing by an item's proper noun status is plotted in Figure 5. /θ/-final items are not included as no /θ/-final items appeared in the data as proper nouns.

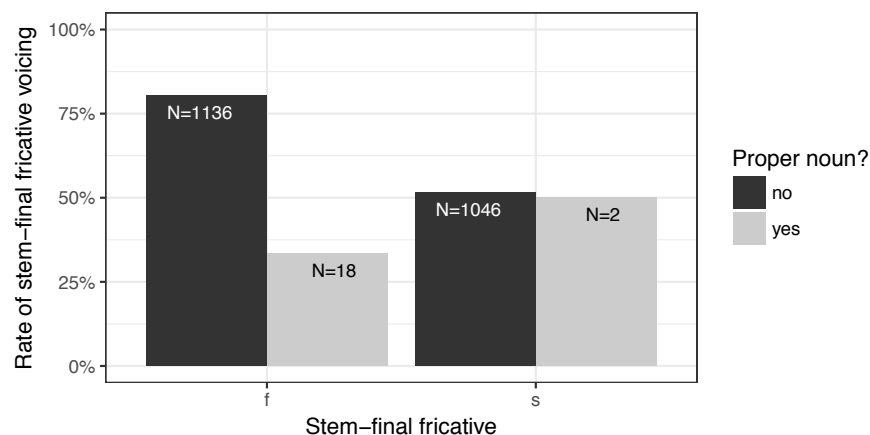


Figure 5: Rate of stem-final fricative voicing by identity of stem-final fricative and item's proper

noun status.

Token counts are very imbalanced; proper nouns are rare in the data (20 tokens out of a total of 2,361, and all but two are /f/-final). However, proper noun status is a significant, negative predictor of stem-final fricative voicing in the pooled data ($\beta = -1.65$, $p = 0.004$) and in the /f/-final subset ($\beta = -2.059$, $p = 0.001$).

The proper nouns ending in /f/ in this data are primarily instances of three lexicalized forms: Snow White and the Seven Dwarfs (a movie title which, in its published form, is not spelled with stem-final fricative voicing), Maple Leafs (the Toronto hockey team, which also has no stem-final fricative voicing in its official spelling), and Timberwolves (the Minnesota basketball team, which **does** have stem-final fricative voicing in its official spelling). Despite the spelling of Timberwolves, however, 80% of tokens of that team name were produced without stem-final fricative voicing ($N = 5$). Pinker's proposal appears to be upheld, then, even when failure to apply irregular morphology contrasts with an official codified spelling.

4.6 Voicing of following segment

Figure 6 examines whether there is less audible voicing of fricatives that precede a voiceless segment or a pause.

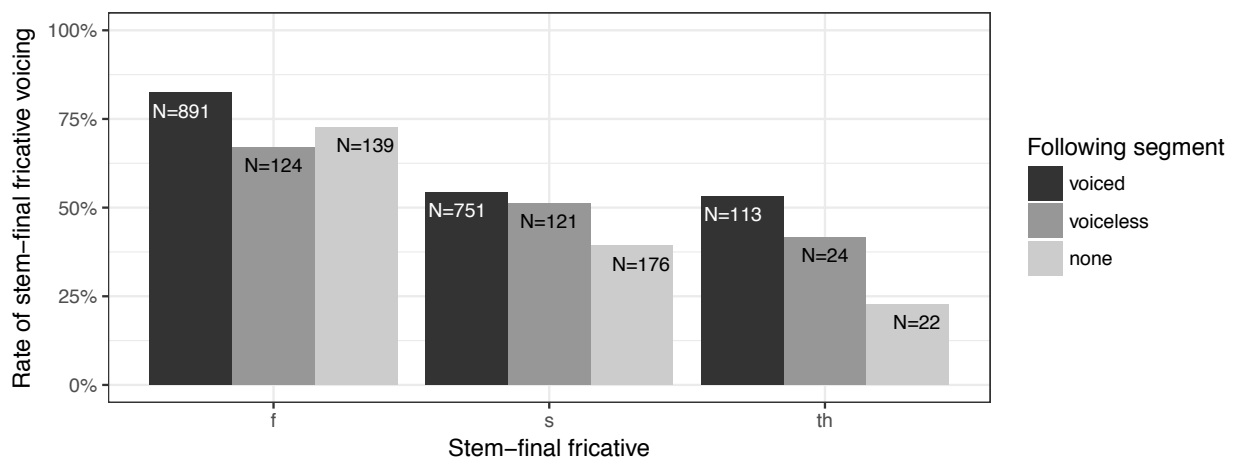


Figure 6: Rate of stem-final fricative voicing by identity of stem-final fricative and voicing of following segment.

Indeed, in the pooled data, we find negative effects on stem-final fricative voicing of a following voiceless segment ($\beta = -0.582$, $p = 1.12 \times 10^{-4}$) and a following pause ($\beta = -0.411$, $p = 0.003$): just as predicted, stem-final fricative voicing is heard less often in these two environments compared to when a plural noun precedes a voiced segment.

This same pattern is found in the /f/-final subset (effect of a following voiceless segment compared to a following voiced one: $\beta = -1.234$, $p = 1.5 \times 10^{-7}$; effect of no following segment: $\beta = -0.488$, $p = 0.029$). In the /θ/-final subset, we also find a negative effect on stem-final fricative voicing of a following pause compared to a following voiced segment ($\beta = -1.331$, $p = 0.016$), though the environment of a following voiceless segment does not reach significance; this is consistent with Smith's (1997) finding that phrase-final is a stronger environment for devoicing than pre-voiceless consonant. Surprisingly, the effect of phrase-finality on stem-final fricative voicing is even attested in houses and its derivatives ($\beta = -0.377$, $p = 0.043$), where assimilation of the stem-final fricative to a devoiced plural suffix is not expected to occur, given the intervening vowel. The effect is weak, though; it could be representative of rare cases in which devoicing extends all the way through the vowel of the plural suffix to the stem-final fricative.

Given this, it appears that phrase-final and pre-voiceless contexts **underestimate** the rate of application of stem-final fricative voicing. We must assume that some of the tokens in these contexts that were heard by coders as voiceless actually had stem-final fricative voicing apply to them, but that this voicing was effectively “undone” by the devoicing of final /z/ and the spreading of this voicelessness to the stem-final fricative. By contrast, because there is no evidence from the literature that a **voiced** word-initial consonant can cause a **voiceless** consonant that pre-

cedes it to voice (Jansen 2007; Myers 2010), we have no reason to believe that any tokens in pre-voiced environments that were coded as having stem-final fricative voicing actually did not have that voicing underlyingly and only gained it via assimilation. Pre-voiced tokens thus give us the most accurate representation of the rate of stem-final fricative voicing in this data. However, because omitting all tokens in other environments would restrict the data too severely, I retain them, but include the important predictor of following segment voicing when modeling.

4.7 Paradigm uniformity

Figure 7 shows the rate of stem-final fricative voicing by the identity of the stem-final fricative and whether a lexeme has a voiced fricative-final counterpart in the lexicon or not. This predictor tests whether, among /f/-final lexemes, the plurals of calf, half, life, shelf, and thief show stem-final fricative voicing at a higher rate than the plurals of beef, dwarf, elf, hoof, knife, leaf, loaf, roof, scarf, wife, or wolf; and, among /θ/-final lexemes, whether the plurals of bath, cloth, mouth, and wreath show stem-final fricative voicing at a higher rate than the plurals of moth, oath, path, truth, or youth. /s/-final items are not plotted, as they represent a single lexeme (house) for which there is a voiced counterpart.

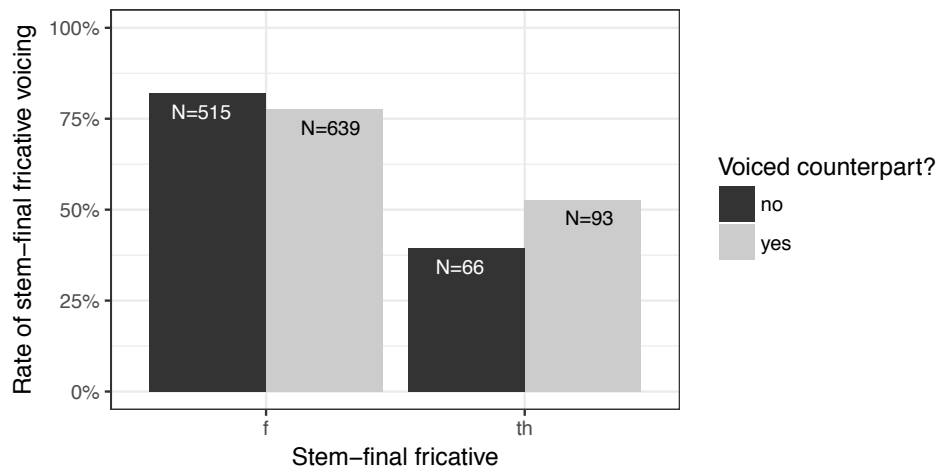


Figure 7: Rate of stem-final fricative voicing by identity of stem-final fricative and presence of

voiced fricative-final counterpart in the lexicon.

Though it looks like this factor could be having the predicted effect for /θ/-final items, it turns out not to be a significant predictor for that subset of data, nor for /f/-final items or for the data pooled across the two stem-final fricatives. There is no evidence that paradigm uniformity plays a role in stem-final fricative voicing variation.

4.8 Year of birth

Figure 8 shows the rate of stem-final fricative voicing by the identity of the stem-final fricative and a speaker's year of birth.

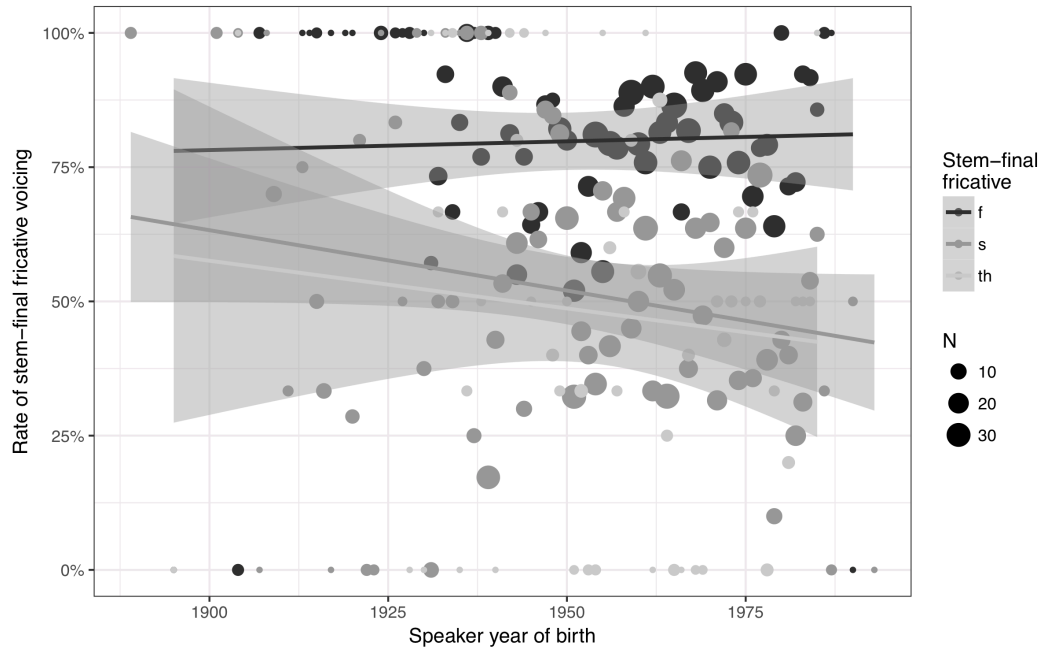


Figure 8: Rate of stem-final fricative voicing by identity of stem-final fricative and speaker's year of birth. Trend lines represent generalized linear smooths with 95% confidence intervals shaded in gray. Dot size reflects the number of tokens from that birth year + stem-final fricative combination in the data.

We see, first of all, that the trend line for the /f/-final subset is effectively flat. The trend line

for the /s/-final subset is pitched downward, reflecting less stem-final fricative voicing of houses and its derivatives among speakers born more recently. The trend line for the /θ/-final subset appears nearly parallel to that for the /s/-final subset, but the shaded confidence interval is very wide in the early years, and could just as well accommodate a flat line.

In the pooled data, speaker year of birth is a significant, negative predictor, with a β of -0.157 (reflecting the change in log-odds of stem-final fricative voicing with each decade of speakers born since 1960) and a p -value of 2×10^{-6} . This holds even when we restrict the data to only those tokens collected in the period from 1991–2004, when the bulk of the data in this study was collected, giving a more accurate picture of the apparent-time trend ($\beta = -0.178$, $p = 1.8 \times 10^{-5}$).

However, this apparent-time change does not reach significance in the /f/-final or /θ/-final subsets on their own. It is only in houses and its derivatives that the change is concentrated ($\beta = -0.237$, $p = 10^{-7}$).¹¹

What we see, then, is that the historical stem-final fricative voicing rule has fragmented. It remains robust for /f/-final lexemes. It applies at only a low rate — and, as far as this data shows, a stable rate — for /θ/-final lexemes. And in houses and its derivatives, it is being lost in apparent time. Young speakers' voicing rate for the plural of house is much more in keeping with their rate for the plurals of bath or mouth than their (near-categorical) rate for the plurals of wolf or life.

4.9 External factors: education, corpus, sex

Two external factors achieve significance in the pooled data, but their effects turn out to be driven by the houses subset. These are a speaker's education level (coded as less than college vs. college or more), and the corpus from which data was collected (Switchboard, Fisher, or the Philadelphia Neighborhood Corpus). Both are plotted in Figure 9 (education on the left, corpus on the

right; speakers whose education level was unknown or otherwise not coded have been omitted from the leftmost figure).

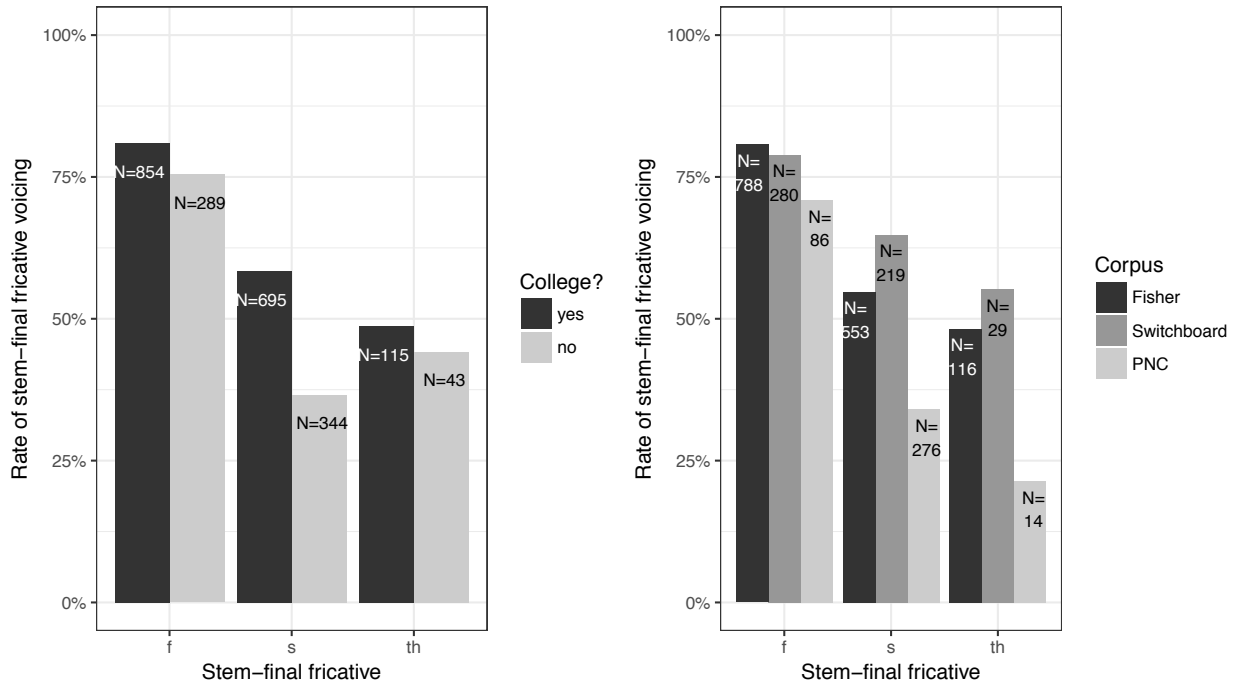


Figure 9: Rate of stem-final fricative voicing by identity of stem-final fricative, speaker's education level (left), and corpus (right).

In both the pooled data and the houses subset, stem-final fricative voicing is disfavored by speakers who have no college (pooled data: $\beta = -0.481$, $p = 2.8 \times 10^{-5}$; houses: $\beta = -0.736$, $p = 3 \times 10^{-6}$) and by speakers in the PNC (pooled data: $\beta = -1.016$, $p = 3 \times 10^{-10}$; houses: $\beta = -1.084$, $p = 3 \times 10^{-8}$).

Recall that the houses subset is the only phonological subset which shows change in apparent time, away from stem-final fricative voicing (Section 4.8). What these results indicate is that this change is led by less educated speakers, and is similarly more advanced among speakers in the PNC. Taking education as a proxy for social class, and bearing in mind that the at-home, socio-linguistic interviews of the PNC are a less formal setting than the between-strangers, assigned-

topic, over-the-phone conversations of Switchboard and Fisher, this is what we'd expect from a change from below: led by speakers from lower social classes, and with the innovative, non-standard variant used more in informal speech (Labov 2001).

Given this, then, it is surprising that there is not a concomitant strong effect in the houses data of speaker sex. Figure 10 plots the rate of stem-final fricative voicing by stem-final fricative and sex.

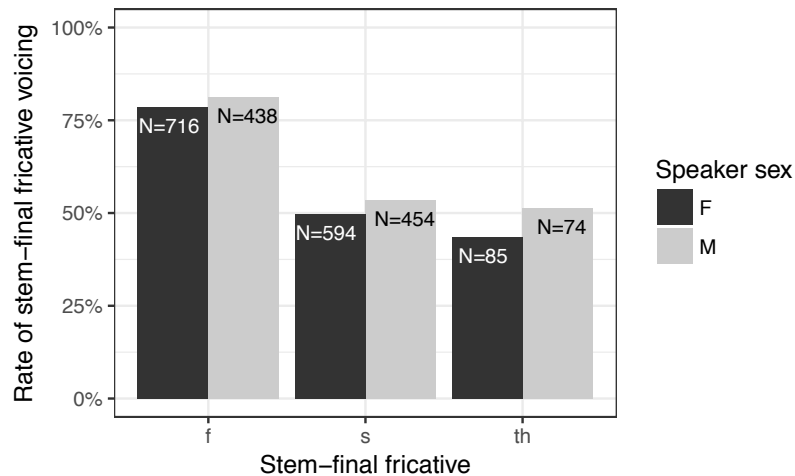


Figure 10: Rate of stem-final fricative voicing by identity of stem-final fricative and speaker sex.

There appears to be a slight effect of more stem-final fricative voicing among men, and indeed this does surface in the regression on the pooled data ($\beta = 0.22$, $p = 0.028$). But it doesn't reach significance in any of the three phonological subsets; the closest it gets is a borderline effect in /f/-final items ($\beta = 0.326$, $p = 0.056$). Still, more use of the innovative variant (that is, non-application of voicing) by women is what we'd expect in the case of a change in progress away from stem-final fricative voicing, even if the effect isn't demonstrated in the subset of data where change is most robust.

4.10 Summary

This paper started out with three research questions, which I answer here by summarizing the above results.

1. Can we confirm previous anecdotal reports that stem-final fricative voicing is variable?
2. If yes, what social and linguistic factors condition the variation?
3. And is this variation part of ongoing change?

The answer to the first question is yes. This section has demonstrated that stem-final fricative voicing is now variable in all three phonological subsets of lexemes whose plurals historically underwent it: /f/-final lexemes, as in wolf; the /s/-final lexeme house; and /θ/-final lexemes, as in bath. Some lexemes were already attested by Jespersen as being variable in their voicing (e.g. roof, path, youth), but we additionally see present-day variability in a number of lexemes for which Jespersen does not mention it (e.g. shelf, house). This study is the first to provide empirical data on variation in speakers' production of the plural forms of these items.

There is an additional question that demands further investigation, concerning whether stem-final fricative voicing has expanded to any lexemes which did not traditionally undergo it, as indicated anecdotally by Hayes (2011) and Becker et al. (2012). Indeed, I note that a voiced token of psychopaths was observed in the data for this study (though it was omitted from analysis as per Section 3.2). A systematic study of stem-final fricative voicing extension is an important topic for future work, particularly given the historical evidence that the set of lexemes that undergo the voicing alternation has both expanded and contracted over time (Section 2.1).

Concerning the second and third questions, a major factor conditioning variability in stem-final fricative voicing is the identity of the stem-final fricative. /f/-final lexemes and /θ/-final lexemes show stably variable voicing in apparent time, though at different overall rates: /f/-final lexemes show voicing at a relatively high rate; /θ/-final lexemes at a relatively low rate. This in-

dicates that analogical leveling of /θ/-final plurals must have started some time ago, while any concomitant change in /f/-final plurals is still only incipient. Change in the plural of the /s/-final house, by contrast, is robust; stem-final voicing in this lexeme drops from a rate of 65% among speakers born in the 1940s (N = 139) to a rate of 38% among speakers born in the 1980s (N = 92).

Additional factors are at play. As has been documented in other studies of speakers' production of irregular items (Hooper 1976; Bybee and Slobin 1982; Yang 2002), frequency plays a role in the analogical leveling of stem-final fricative voicing. The difference in overall voicing rate between /f/-final lexemes and /θ/-final lexemes correlates with the different relative frequencies of those two phonological classes. Moreover, within each of the /f/-final and /θ/-final subsets, we find voicing to correlate with plural form frequency. Proper nouns and opaque compounds additionally disfavor stem-final fricative voicing (an effect that manifests only for /f/-final lexemes for reasons of data sparsity), and there is less voicing before a pause if not also before a voiceless segment, presumably due to an independent effect of voicelessness assimilation that may obscure the application of stem-final fricative voicing. Finally, where change is most robust — in the /s/-final lexeme house — we find traditional hallmarks of a change from below (Labov 2001): less voicing among less-educated speakers and in the more casual sociolinguistic interviews of the PNC as compared to the corpora of telephone conversations.

5 Discussion

The present study has found evidence for phonological class behavior in both variation and change in stem-final fricative voicing: /f/-final lexemes are consistent in stably voicing at a high rate, /θ/-final lexemes in stably voicing at a low rate, and houses and its derivatives are changing over time. The tendency for irregular forms to show class behavior is well attested, both in lan-

guage processing (Regel et al. 2015) and in first language acquisition (Yang 2002). In this section, I suggest that the patterns in the data presented here are reminiscent of how children learn irregular forms, and that these adult language patterns may have their source partly in child language acquisition errors.

Yang (2002:79) observes class behavior in native English-speaking children's successful acquisition of English irregular past tense forms. He shows that irregular English past tenses can be grouped into a number of phonologically-defined subsets, or classes: these include verbs for which the vowel of the rhyme changes to /ɔ/ and the coda changes to /t/ (e.g. catch–caught, think–thought, bring–brought, buy–bought); verbs for which the past tense is zero-inflected (e.g. put, hit, hurt, cut); verbs for which the vowel of the rhyme backs but no other change occurs (e.g. get–got, take–took, write–wrote, win–won), and so forth. Yang shows that, within a given class, the more frequent a verb is in the adult's input to the child, the more accurately the child learns the irregular form. Less frequent verbs are less likely to be learned in their irregular form and more likely to be regularized by children. Crucially, however, this frequency behavior is only visible on a class level. Ordering all irregular verbs from most to least frequent does not show a cleanly decreasing effect of learning accuracy. It is only when verbs are grouped into their phonological classes and **then** sorted from most to least frequent within those classes that Yang finds a clear correlation with learning accuracy.

The reason that simply ordering all irregular verbs from most to least frequent does not cleanly correlate with learning accuracy is what Yang calls “the free-rider effect”: the importance of the overall frequency of a class in boosting the learning accuracy of the low-frequency members of that class. An infrequent item can still be learned with a relatively high rate of accuracy if it belongs to the same class as one or more very high-frequency items. For

instance, the past tense forms hurt and drew occur at comparable frequencies in adult language, yet children achieve much higher accuracy on hurt than on drew (Yang 2002:81). Yang argues that this can be attributed to hurt's membership in a class which also contains the high-frequency hit, let, set, cut, and put: hurt gets a “free ride” learnability boost from those high-frequency forms. The past tense form drew, by contrast, has no such high-frequency counterparts in its class, so it receives no boost.

We can see similar patterns at play in the present study. Figure 3 shows a number of /f/-final lexemes (black dots) whose frequency in their plural form is roughly equivalent to that of one or more /θ/-final lexemes (lightest gray dots). However, in nearly all cases, the /f/-final lexeme achieves a higher rate of stem-final fricative voicing than the /θ/-final lexeme. Compare, for instance, loaf (Zipf scale frequency of plural form: 2.77; voicing rate: 50%), scarf (Zipf scale frequency of plural form: 2.83; voicing rate: 74%), and youth (Zipf scale frequency of plural form: 2.85; voicing rate: 0%). The three have comparable plural frequencies, but we see much higher voicing rates for the two /f/-final lexemes than the /θ/-final lexeme. This may reflect a “voicing boost” coming from high-frequency, high-voicing /f/-final lexemes like life (Zipf scale frequency of plural form: 5.15; voicing rate: 80%) and leaf (Zipf scale frequency of plural form: 4.66; voicing rate: 95%), not to mention the anaphoric -selves pronouns, which were omitted from the present study, but are highly frequent and were found in pilot work to voice at a high rate of 84% (N = 128). This suggests a role for first-language acquisition processes in synchronic patterns of stem-final fricative voicing variation in adult speech. This irregular voicing alternation may be more accurately learned for lexemes whose plural form frequency is high, as well as for items in the same phonological class as them — i.e., lexemes that end in the same fricative. Lexemes whose plural form has a lower absolute frequency, and those in the same phonological class as

them, show more regularization.

As indicated in Section 4.3, the frequency/irregularity correlation isn't perfect. A clear exception is seen for house, whose plural form has the third-highest frequency in this study (Zipf score: 4.28), but which is rapidly losing voicing in apparent time. A possible reason for this is the lack of other /s/-final fricative voicing lexemes to support house: in a class of its own, its high frequency is not enough to maintain its irregularity. Indeed, Yang (2002:82) finds exactly this sort of pattern in the production of irregular past tense forms by “Abe,” a child in the CHILDES corpus. Despite the high frequency of go and come, Abe performs poorly at producing the past tenses of these verbs, which Yang attributes entirely to their being the sole members of their respective classes.

All this being said, there is an additional conservative effect operating on the plurals of /f/-final lexemes: orthography. The /f/-final items are the only ones in the set in (2) which change their spelling in the plural (from <f> to <v>). So, another possible explanation for the patterns presented here is that there is a general trend toward across-the-board leveling of stem-final fricative voicing, well underway for /θ/-final lexemes, and rapidly progressing for house, but blocked (or at least strongly slowed) in /f/-final lexemes because of the standardizing influence of their spelling. If accurate, this would be an interesting and perhaps novel case of orthography influencing the trajectory of language change.¹²

Because orthography and frequency are so strongly connected in this study — six out of the the top seven most frequent lexemes change their spelling in the plural — it is important to tease apart their effects. To this end, a direction for follow-up research is to investigate stem-final fricative voicing among preliterate children. Do young children already voice /f/-final plurals at a high rate due to the frequency boost they receive for these forms in the input? Or do they regular-

ize /f/-final plurals at a level comparable to that shown for /θ/-final and /s/-final plurals, only to grow out of this after learning how they are standardly spelled? Studies of child speech corpora and elicitation tasks with speakers of different ages will be welcome means of resolving this issue.

6 Conclusion

This paper has provided the first corpus-based study of stem-final fricative voicing in a small set of English plural nouns (e.g. houses [haʊzəz], leaves [livz], paths [pæðz]). I have demonstrated that stem-final fricative voicing in the set of lexemes considered here is variable, and shows clear phonological class behavior. /f/-final lexemes voice at a high, stable rate in apparent time; /θ/-final lexemes at a low, stable rate in apparent time, and the /s/-final lexeme house is rapidly losing voicing in apparent time. The frequency of a lexeme in its plural form and the overall frequency of the phonological class it belongs to both play significant roles in the variation, with more frequent plurals and lexemes in the more frequent /f/-final class resisting leveling. This pattern is reminiscent of the nature of the errors children produce as they acquire the English past tense. As such, it suggests that the source of the variation may lie in children's first-language acquisition errors. More generally, it underscores that a careful consideration of the morphophonological conditions on variation can enrich our understanding of a variable phenomenon.

The connection between child language acquisition errors and diachronic directions of change has been observed before in other studies of analogical leveling (Hooper 1976), though the present study's specific demonstration of child language-like phonological class behavior in adults' probabilistic production is, to my knowledge, new. Truly connecting the dots between children's regularization errors and the variation and change in stem-final fricative voicing documented here will require case studies of children's production of these irregular plural forms,

particularly given the potentially confounding factor of orthography. Since the items studied here are all nouns, many of them quite common and early-learned (e.g. houses, wolves, baths), this should not be particularly difficult to do. Future work that expands on the results presented here will be able to shed even more light on the connection between children’s acquisition errors and patterns of language variation and change.

¹ In the fifteenth century, the /-əs/ plural suffix changed to /-z/ (Jespersen 1909), but stem-final fricative voicing was nonetheless maintained in these forms. This provides further evidence that a regular intervocalic voicing rule was no longer the source of their voicing.

² Jespersen also includes the obscure delf ‘that which is delved’ and loof ‘palm of the hand’. Additionally, he mentions two words for which a voiced fricative is attested in the plural, but also in the singular: belief (“formerly [singular] -ve, [plural] -ves”) and sheriff (“the M[iddle] E[nglish] form had /v/ also in the [singular]”). These were omitted from study because the historical singular [v] means that any [v] in the plural today could have a different historical source than that of the other words studied here.

³ With the exception of the 171 tokens of plural leaf, which were coded by the author.

⁴ It should be pointed out that the student coders were native speakers of British English, coding data from American English speakers. There is no reason to think that this should necessarily affect their ability to accurately code the voiced vs. voiceless distinction; this distinction exists in British English as it does in American English, and, though the precise phonetic means by which it is implemented in American English is complicated (see next paragraph), this is the case for British English as well (see literature cited in Davidson 2016).

⁵ In fact, Fruehwald (2017) addresses this same question using data from five different language changes in the PNC: are the observed age effects the result of age-grading or of generational change? He finds generational change to be much more common across the five changes than age-grading. Fruehwald recommends using generalized additive mixed models with two-dimensional tensor product smooths to analyze real-time data like this, since they can simultaneously assess effects of generational change and age-grading; however, he observes that this modeling method has difficulty handling additional fixed effects. The large number of fixed-effect predictors that needed to be considered in the present study made this type of modeling untenable.

⁶ Though the Switchboard manual provides no clarity on this point, I assume that the corpus has coded speakers who were in college at the time of recording as “college,” rather than “less than college.” This is because Switchboard’s “less than college” category contains no speakers younger than 22 in this data, which one would expect it to had college students been coded as such.

⁷ The Switchboard data contains no speakers under 20, but teenagers are represented in both the PNC and Fisher. An important question is how many “less than college” speakers from Fisher and the PNC were in fact not old enough to have attended college yet, or were in their first year of college (and accordingly reported their years of schooling as 12). If it is a large proportion, we have a serious confound between education level and age. However, only 7% of Fisher/PNC

speakers who reported 12 or fewer years of schooling are 19 or under, indicating that the large majority of these speakers are old enough to have attained more schooling.

⁸ Final model formula, in R's lme4 syntax: `Voicing ~ Stem-final.fricative + Morphological.complexity + Following.segment.voicing + Plural.stem.frequency + Proper.noun + Sex + Education + Corpus + Year.of.birth + (1|item)`.

⁹ So do elf and beef, but they are each only represented in this data by a single token.

¹⁰ Separate frequencies could not be obtained for the different meanings of a polysemous pair (e.g. calf=animal vs. calf=muscle), so polysemous lexemes are represented by a single dot here.

¹¹ Again, this result remains robust even when the data is restricted to those tokens collected in the period from 1991–2004 ($\beta = -0.322$, $p = 10^{-7}$).

¹² De Schryver et al. (2008) propose a similar explanation for patterns of variation in their study of fricative devoicing in Dutch, though rather than studying speakers' pronunciations of the fricatives in question, they asked subjects to choose which of two spellings of the fricative they thought was the correct one for a series of words. Their results may thus exaggerate the influence of orthography compared to what might be found in a production study.

References

- Bates, Douglas, Martin Mächler, Ben Bolker and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48.
- Becker, Michael, Andrew Nevins and Jonathan Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. Language 88:231–268.
- Boersma, Paul and David Weenink. 2015. Praat: Doing phonetics by computer [computer program].
- Brysbaert, Marc and Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods 41:977–990.
- Bybee, Joan L. and Dan I. Slobin. 1982. Rules and schemas in the development and use of the English past tense. Language 58:265–289.
- Chambers, J.K. 2009. Sociolinguistic theory: Linguistic variation and its social significance. Oxford: Wiley-Blackwell.
- Chomsky, Noam and Morris Halle. 1968. The sound pattern of English. Cambridge, MA: The MIT Press.
- Cieri, Christopher et al. 2004. Fisher English training speech parts 1 and 2. Philadelphia: Linguistic Data Consortium.
- Clopper, Cynthia G. 2011. Checking for reliability. In Marianna Di Paolo and Malcah Yaeger-Dror (eds.), Sociophonetics: A student's guide, 188–197. London: Routledge.
- Davidson, Lisa. 2016. Variability in the implementation of voicing in American English obstruents. Journal of Phonetics 54:35–50.

- De Schryver, Johan, Anneke Neijt, Pol Ghesquière, and Mirjam Ernestus. 2008. Analogy, frequency, and sound change: The case of Dutch devoicing. Journal of Germanic Linguistics 20:159–195.
- Fruehwald, Josef. 2017. Generations, lifespans, and the zeitgeist. Language Variation and Change 29:1–27.
- Godfrey, John J. and Edward Holliman. 1997. Switchboard-1 release 2. Philadelphia: Linguistic Data Consortium.
- Hayes, Bruce. 2011. Introductory phonology. Malden, MA: Blackwell.
- Hooper, Joan B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In Current progress in historical linguistics: Proceedings of the Second International Conference on Historical Linguistics, Tucson, Arizona, 12-16 January 1976, 96–105. Amsterdam: North Holland.
- Jansen, Wouter. 2007. Phonological voicing, phonetic voicing, and assimilation in English. Language Sciences 29:270–293.
- Jespersen, Otto. 1909. A modern English grammar on historical principles, part I: Sounds and spellings. Heidelberg: Carl Winters Universitätsbuchhandlung.
- Jespersen, Otto. 1942. A modern English grammar on historical principles, part VI: Morphology. Copenhagen: Ejnar Munksgaard.
- Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. Language and Linguistics Compass 3:359–383.
- Labov, William. 2001. Principles of Linguistic Change: Social Factors. Malden, MA: Blackwell.
- Labov, William and Ingrid Rosenfelder. 2011. The Philadelphia Neighborhood Corpus.

- Lass, Roger. 2000. Phonology and morphology. In Roger Lass (ed.), The Cambridge history of the English language volume 3: 1476–1776, 56–186. Cambridge: Cambridge University Press.
- Myers, Scott. 2010. Regressive voicing assimilation: Production and perception studies. Journal of the International Phonetic Association 40:163–179.
- Phillips, Betty S. 1984. Word frequency and the actuation of sound change. Language 60:320–342.
- Pinker, Steven. 1994. The language instinct. New York, NY: William Morrow; Company.
- R Core Team. 2016. R: A language and environment for statistical computing.
- Raphael, Lawrence J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. The Journal of the Acoustical Society of America 51:1296–1303.
- Regel, Stefanie, Andreas Opitz, Gereon Müller and Angela D. Friederici. 2015. The past tense debate revisited: Electrophysiological evidence for subregularities of irregular verb inflection. Journal of Cognitive Neuroscience 27:1870–1885.
- Ringe, Donald and Joe Eska. 2013. Historical linguistics: Toward a twenty-first century reintegration. New York: Cambridge University Press.
- Sankoff, Gillian. 2005. Cross-sectional and longitudinal studies in sociolinguistics. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier and Peter Trudgill (eds.), Sociolinguistics: An international handbook of the science of language and society, vol. 2, 1003–1012. Berlin: Walter de Gruyter.
- Seyfarth, Scott. 2016. Contextual and morphological effects in speech production. University of California, San Diego PhD thesis.

- Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. Journal of Phonetics 25:471–500.
- van Heuven, Walter J. B., Pawel Mandera, Emmanuel Keuleers and Marc Brysbaert. 2014. SUBTLEX-UK: A new and improved word frequency database for British English. Quarterly Journal of Experimental Psychology 67:1176–1190.
- Weide, Robert. 2008. The CMU Pronouncing Dictionary.
- Yang, Charles. 2002. Knowledge and learning in natural language. New York: Oxford University Press.