DE GRUYTER MOUTON
DOI 10.1515/cllt-2013-0025 — Corpus Linguistics and Ling. Theory 2014; 10(2): 213–241

Earl K. Brown*, Michael S. Gradoville and Richard J. File-Muriel

The variable effect of form and lemma frequencies on phonetic variation: Evidence from /s/ realization in two varieties of Colombian Spanish

Abstract: Research has shown that frequency conditions the variable realization of sounds. However, the literature has not addressed whether the frequency of the individual word forms, or form frequency, has a larger conditioning effect than the combined frequencies of the members of the paradigm to which the forms belong, or lemma frequency. Monofactorial correlation tests and monofactorial and multifactorial linear regression analyses are performed on 2,734 tokens of Spanish /s/ in sociolinguistic interviews conducted in Cali and Barranquilla, Colombia. Two findings are highlighted: (1) frequency is only significant in the variety of Spanish that has low overall rates of /s/ reduction, Cali, and (2) form frequency is more influential than lemma frequency.

Keywords: frequency effects, form frequency, lemma frequency, Spanish /s/, Colombian Spanish, activation state

***Corresponding author: Earl K. Brown:** Kansas State University. E-mail: ekbrown@ksu.edu
Michael S. Gradoville: Spelman College. E-mail: mgradovi@spelman.edu
Richard J. File-Muriel: University of New Mexico. E-mail: RichFile@unm.edu

1 Introduction

Frequency has been shown to condition sound reduction in speech. In general, more frequently used words exhibit reduction of sounds while less frequently used words maintain sounds more robustly. This phenomenon was noted at least as long ago as the late nineteenth century. Schuchardt (1885) (reprinted and translated in Venneman and Wilbur 1972: 58) proposes that, with regard to sound changes, "rarely-used words drag behind; very frequently used ones hurry ahead." Modern-day examples abound. Post-tonic schwa is more likely to be deleted in more frequent words in English, such as in the more frequently used words *memory* and *nursery* than in less frequent *armory* and *cursory* (Hooper

1976). More frequent multi-word strings with English *don't* are more likely to be reduced than less frequent ones, for example the more frequent *I don't know* (to *I dunno*) in comparison to the less frequent *I don't believe* (Bybee and Scheibman 1999; Scheibman 2000). More recent examples come from the phonetic and phonological reduction of Spanish /s/. Brown and Torres Cacoullos (2003) demonstrate a reductive effect from frequency in the speech of four male speakers in Chihuahua, Mexico. Minnick Fox (2006) shows a conditioning effect from frequency in fifteen Latin American varieties of Spanish. File-Muriel (2009) even proposes that frequency is the most influential factor in the realization of word-medial, syllable-final lexical /s/ in Barranquilla, Colombia. Studying another variety of Colombian Spanish, Brown (2009a) finds a significant effect from frequency in syllable- and word-final /s/ in Cali. In another corpus of Cali Spanish, File-Muriel and Brown (2011) find frequency as significant in the realization of both syllable-initial and -final /s/. Measuring and controlling for frequency in the variable realization of sounds has been well documented in the literature. The few studies mentioned here are only a brief overview of the many that could have been cited (such as Bush 2001; Alba 2006; Eddington 2011; Diaz-Campos and Gradoville 2011; Diaz-Campos et al. 2012, among others).

It should be noted that, as proposed by Phillips (1984), more frequent words are affected by sound changes first and to a greater degree when those changes are reductive in nature. However, when a given sound change is not reductive in nature, it may be that high-frequency words are not necessarily affected more than low-frequency words. Dinkin (2008) finds this to be the case in the shift of short vowels in North American English labeled the Northern Cities Shift.

In many of the studies that find a conditioning effect from frequency, including the few cited above, the frequency of individual forms was measured rather than the frequency of the lemmas to which individual words belonged. In other words, form frequency has been the focus rather than lemma frequency.

Notwithstanding the prevalence of form frequency as the mechanism of measurement in many studies, several studies analyze the influence of lemma frequency, but report conflicting results about its effect. Gahl (2008) finds that homophones in English, such as *time* and *thyme*, are not completely homophonous in phonetic terms and that tokens belonging to higher-frequency lemmas, such as *time*, are more likely to have a shorter duration than tokens belonging to lower-frequency lemmas, such as *thyme*. Interestingly, Jurafsky et al. (2002) do not find an effect from lemma frequency even though their data come from the same corpus that Gahl used to study homophones (the Switchboard Corpus). Likewise, Jescheniak and Levelt (1994) fail to find an effect from lemma frequency in other production data. Overall, the literature does not agree on the

effect of lemma frequency in production data.[1] Clearly, further research is needed to elucidate the influence of lemma frequency on linguistic variation, if one exists.

In addition to the discussion regarding form and lemma frequencies, several recent studies suggest a more complex relationship between frequency on the one hand and phonological and syntactic variation on the other. For example, Walker (2012) analyzes (t/d)-deletion impressionistically and suggests that frequency may not operate monotonically, but rather has a more dynamic interaction with the lexicon: "The usage-based hypothesis that lexical frequency constrains (t/d)-deletion receives no support, once its interaction with morphological status and the contribution of a small set of lexical items are taken into account" (p. 412). Regarding the expression of Subject Personal Pronouns (SPP) in Spanish, Erker and Guy (2012) find that high frequency does not by itself favor or disfavor pronoun use, but rather has significant interactions with other constraints affecting SPP occurrence: "High frequency potentiates or amplifies the effects of other constraints" (p. 527). At the same time, Bayley et al. (2013) present evidence that frequency has neither a magnifying nor activating effect on SPP use as indicated by Erker and Guy's (2012) study. Rather, frequency has a relatively small effect on SPP use among speakers of Mexican-descent in the San Francisco Bay area, California and in San Antonio, Texas.

Clearly, the manner in which lexical frequency operates in language variation remains controversial. This paper seeks to contribute to our understanding of the influence of frequency on variation by comparing the conditioning effect of form frequency and that of lemma frequency on the variable phonetic realization of syllable-initial and -final /s/ in two varieties of Colombian Spanish. One variety, Cali, displays low rates overall of /s/ reduction while the other, Barranquilla, shows high rates in syllable-final position. The purpose of this paper is to determine which frequency measurement is more influential on this linguistic variable.

The theoretical approach of this paper is usage-based, that is, it is assumed that "grammar is the cognitive organization of one's experience with language" (Bybee 2006: 711). It is based on exemplar theory (Nosofsky 1988), as exemplar models are well suited to account for frequency effects, including differences in frequency among words and phrases. It is assumed that "words hav[e] a memory representation that [consists of] a phonetic range, that is, a cluster of

1 The apparently contradictory effect of lemma frequency in production and perception, that is, that lemma frequency is more important than form frequency in perception, as proposed by Jurafsky et al. (2002: 21), is out of the scope of this paper and should be taken up elsewhere.

exemplars . . . , rather than an abstract phonemic representation" (Bybee 2010: 20). Additionally, it is proposed that frequency information about individual words is stored in the exemplar cluster of that word. That is, form frequency is believed to be stored with the mental representation of the word. Further, the notion that lemma frequency attempts to describe can be conceived of as the combined influence of the frequency of individual members of an inflectional paradigm, as transmitted through associative network connections (Bybee 1985). These associative networks propose that "a word, which consists of a cluster of phonetic exemplars as well as a set of semantic exemplars, can be considered a unit which can then be related to other words in various ways. Words form relations along phonetic dimensions . . . as well as along semantic dimensions" (Bybee 2010: 22). It is through these relations with other words in an inflectional paradigm that lemma frequency is manifested.

The research questions motivating this paper are:

1. Is form frequency more influential than lemma frequency on the variable realization of /s/ in two varieties of Colombian Spanish, or is the opposite true? In theoretical terms, is the frequency information stored in the exemplar cluster of an individual word more influential than the effect that the frequency information stored with other members in an inflectional paradigm, as transmitted through associative networks?
2. Is /s/ affected by frequency in the same manner in the two varieties of Colombian Spanish, one with a low overall rate of /s/ reduction and the other with a high level?

As will be seen below, the results of this paper show that form frequency is more influential than lemma frequency, but only in the Cali data; in the Barranquilla data, neither frequency measurement plays a role in the realization of /s/. As such, this paper suggests that the frequency information stored in exemplars of individual word forms is more influential than the influence of the frequency of other members of the inflectional paradigm to which a given word belongs, as transmitted through associative networks.

2 Data and methods

In order to analyze the conditioning effect of the two frequency measurements and determine which one, if either, has a larger conditioning effect, /s/ realization in two varieties of Colombian Spanish is analyzed. One variety, Cali, has relatively low rates of /s/ reduction while the other one, Barranquilla, has higher

rates. It has been proposed that frequency loses its conditioning effect in the (very) advanced stages of a sound change (cf. Labov 1994: 65; Phillips 2006: 13; Brown 2009b: 207). As such, it is hypothesized that frequency will have a smaller effect in Barranquilla than in Cali.

The Spanish of Cali, unlike most varieties of Spanish, exhibits variable /s/ realization in both syllable-initial and -final positions. As such, in the Cali data, /s/ in all word positions is analyzed, that is, word-initial /s/, which is categorically in syllable-initial position (e.g. *ciudad* 'city'), word-medial /s/ in both syllable-initial (e.g. *casi* 'without') and -final (e.g. *especial* 'special') positions, and word-final /s/ in both syllable-initial (e.g. *de vez en cuando* 'once in a while') and -final (e.g. *no es que* 'it's not that') positions are included in the study. The Spanish of Barranquilla patterns with the majority of Spanish varieties in that variable /s/ realization is limited to syllable- and word-final positions. Consequently, in the Barranquilla data, only /s/ in word-medial, syllable-final position and word-final /s/, in both syllable-initial and -final positions, is investigated. While the Spanish of Cali can safely be considered a highland or highland-like Colombian dialect, given its relatively high levels of /s/ reduction in word-initial and word-medial, syllable-initial positions (cf. Lipski 1994: 209), the Spanish of Barranquilla can safely be considered part of the Caribbean macro-dialectal region, which displays virtually no reduction of /s/ in these positions. For this reason, only /s/ in word-medial, syllable-final and word-final positions is analyzed in the Barranquilla data while /s/ in all positions is analyzed in Cali.

The data were extracted from two corpora of sociolinguistic interviews (cf. Labov 1984), one of which was recorded in Cali in 2004 while the other was collected in Barranquilla in 2009. In each case the speakers were native to the city in question as were the interviewers. For this study, small subsets of the total number of speakers in each corpus were studied: eight from the Cali corpus, all of whom were women, and thirteen from the Barranquilla corpus, seven of whom were women while six were men. The speakers were university graduates or students and thus can safely be considered members of the middle to upper-middle class. Socio-cultural class was purposefully controlled for in order to create a relatively homogeneous social group and therefore reduce differences caused by sociolect.

As is the case with sociolinguistic interviews, the topics of the interviews varied according to the interests of the speakers. Topics included vacation and holiday plans, recreation, educational pursuits, food, local shopping, and even dangerous situations that the speakers had experienced. The interviews were conducted in a quiet room with a Marantz PMD 670 solid-state compact flash recorder using a head-mounted unidirectional microphone.

2.1 Dependent variables

The phonetic realization of /s/ in the two varieties of Colombian Spanish was measured by analyzing three acoustic correlates: (1) duration of /s/ in milliseconds, (2) center of gravity of /s/ in Hertz, and (3) percentage of voicelessness of /s/. The measurement of these three dependent variables was performed with the phonetics software Praat (Boersma and Weenink 2012). The decision to perform a fine-grained analysis rather than placing the tokens into bins, such as *maintained* or *reduced*, is based on the fact that nuances about the variable nature of /s/ realization are lost when /s/ is analyzed as a categorical variable. Similarly, when /s/ realization is analyzed by means of various gradient metrics, the variable nature of the predictor factors that condition the realization of /s/ is made apparent. In other words, while a given predictor variable may have a significant effect on one acoustic correlate, it may not significantly condition another acoustic correlate. This type of nuance is lost when /s/ realization is analyzed as a categorical variable (cf. File-Muriel and Brown 2010, 2011; Erker 2010). Consequently, it is necessary to analyze the acoustic correlates individually in order to achieve the purpose of this study, that is, of obtaining an accurate measurement of the conditioning effect of form and lemma frequencies on /s/ realization in all its manifestations, whether it be duration, strength of sibilance as measured by center of gravity, or amount of voicing.

The boundaries of each token of /s/ were manually delimited within a TextGrid in Praat by the first and third authors while viewing both the wave form and the spectrogram of the token. Subsequently, a script automated the extraction of the three dependent variables. The duration was simply calculated based on the manually delimited boundaries. The percentage of voicelessness of /s/ was taken from Praat's Voice Report when the segment filled exactly one-third of the Editor Window. The voicing was taken when the segment filled one-third of the window in order to maintain consistency of measurement across the tokens, as the percentage of voicing given by the Voice Report can fluctuate slightly depending on how much of the segment is visible. This is especially true when less than 5% or more than 95% of the segment is visible.[2] The center of gravity was calculated using only the middle 60% of the manually delimited duration in order to avoid articulatory overlap with the surrounding sounds, which could have occurred

2 While prescriptive grammar (e.g. Schwegler et al. 2010: 304) holds the view that voicing of /s/ can only occur before a voiced consonant (e.g. *mismo* 'same', *es de* 'is from') and, therefore, voicing is simply not an option in other contexts, the results presented in Section 3 show that voicing of /s/ occurs in all contexts: before consonants, vowels and pauses.

had the entire duration been used. Additionally, a Pass Hann filter eliminated the lowest 750 Hertz of the signal of all tokens before calculating the center of gravity, as voicing greatly skews the center of gravity measurement (cf. Gradoville 2011). In acoustic terms, the center of gravity is a weighted average in Hertz of the frequency peaks over a specified duration of a sound.

The total number of tokens in this study is 2,734, of which 1,749 come from the Cali corpus and 985 come from the Barranquilla corpus. Tokens that were abnormally elongated, specifically 200 milliseconds or more, were excluded as they represented a voiced pause that the speakers used as they formulated their next utterance, such as with the words *este* 'this' and *pues* 'well'. Also, in order to avoid the more artificial language that can be common in the first part of an interview, a sequential set of tokens from each speaker was extracted starting at least ten minutes into each interview.

As noted above, Barranquilla Spanish displays higher levels of overall /s/ lenition than Cali Spanish and, as such, it is understandable that there is a higher percentage of deleted tokens of /s/ in Barranquilla than in Cali. Specifically, of the 985 tokens of /s/ in the Barranquilla data 425 tokens, or 43%, are deleted, that is, the duration of these tokens is zero. Differently, of the 1,749 tokens of /s/ in the Cali data, 268 tokens, or only 15%, are deleted. These deleted tokens are assigned a percentage of voicing of 100% and the lowest attested center of gravity of a maintained /s/ in each data set: 920 Hertz in the Barranquilla data and 1309 Hertz in the Cali data. The decision to not assign a center of gravity of zero to the deleted tokens is based on the fact that the surrounding segments themselves have a center of gravity above zero and thus assigning zero would have artificially skewed the data.

2.2 Predictor variables

As discussed above, the purpose of this paper is to gauge the relative influence of form and lemma frequencies on the phonetic realization of /s/. Form frequency was calculated from the C-ORAL-ROM rank-order frequency list (Cresti and Moneglia 2005). This list was employed because it has the lemma of each form listed and, as such, the lemma frequency of the forms could be easily calculated. Moreover, the methodology used to create the list is well documented. To calculate the lemma frequency, the frequency of all forms that belonged to the same lemma were added together. For example, the inflected forms of the lemma *ESTAR* 'to be' were summed together to create a single lemma frequency that applied to all inflected forms of that verb. Consequently, the frequent *está* '3s present indicative to be' and the infrequent *estuviéramos* '1p imperfect

subjunctive to be', for example, have very different form frequencies but have the same lemma frequency because they both belong to the lemma *ESTAR*. It should be noted that, as is the norm in studies dealing with frequency measured as a gradient variable (cf. Gahl 2008: 483–484), the raw frequency numbers were transformed into natural logarithm numbers when entered into the statistical models discussed below. This practice helps control for the outlying most frequent forms and lemmas, as there is always a large disparity in frequency between the most frequent forms and lemmas and all others. Additionally, before fitting the linear regressions, the logged frequency numbers were also transformed to z-standardized numbers (cf. Gries 2013: 130) in order to compare coefficient values between models.

As might be expected, there is a high correlation between form frequency and lemma frequency, as seen in a Kendall's rank correlation tau test ($tau = 0.78$, $p \leq 0.001$). A rank correlation test, which analyzes the data ordinally rather than continuously, was chosen as the assumption of normality among the data that is required to run a Pearson's product-moment correlation test was not met, as evidenced by two Shapiro-Wilk normality tests (form frequency: $W = 0.57$, $p \leq 0.001$; lemma frequency: $W = 0.61$, $p \leq 0.001$).[3] See Figure 1. The diagonal line in the figure marks the word forms whose lemma frequency is equal to their form frequency, that is, word forms that represent the only member of their lemma.

In addition to the frequency measurements, the following, additional, predictor variables were analyzed. These variables have been identified in the literature as contributing to the variable realization of /s/ in Spanish.

The local speaking rate of each token of /s/ was measured as the number of segments per second. This measurement was calculated by taking the duration of the three-word window immediately around /s/, with the /s/ word in the middle, and dividing that duration by the combined number of segments in the three-word string. The /s/ segment itself and its duration were excluded from this calculation in order to avoid circularity in the prediction of /s/. Only the /s/ in question was removed from the calculation, such that words with more than one token of /s/ included the other /s/ in the calculation. When the word with /s/ was immediately preceded or followed by a pause, the number of words in the surrounding window was reduced to two, and occasionally to only the word with /s/ itself. The reason that the local speaking rate of each token of /s/ was calculated is because the rate at which speakers speak naturally fluctuates throughout an interview.

3 It should be noted that, counter-intuitively, a p-value below 0.05 in a Shapiro-Wilk normality test signifies that the data are not distributed normally (cf. Gries 2013: 164).

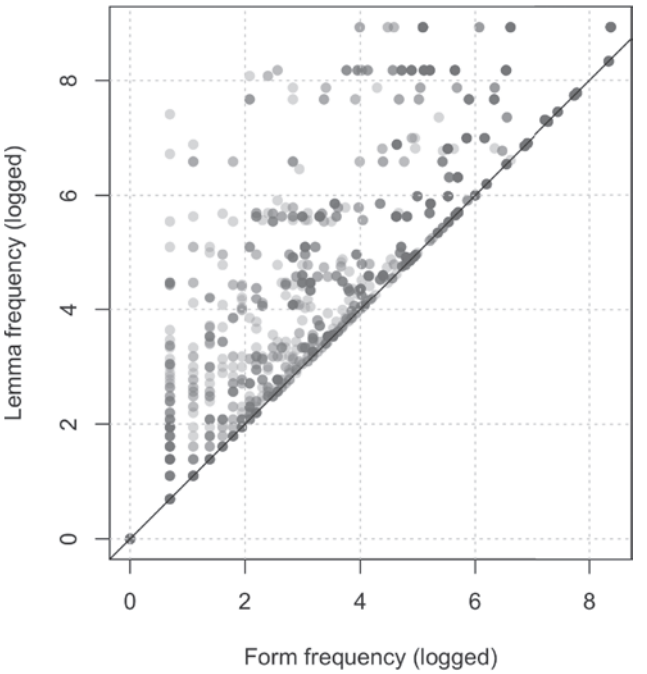

Fig. 1: Correlation between form frequency and lemma frequency

Some topics of discussion may cause a given speaker to speed up because of excitement or fear while other topics may cause speakers to become pensive and therefore speak at a more deliberate pace. Not surprisingly, Gahl (2008) finds a significant main effect from the local speaking rate on the duration of homophones in English.

The length of the words in which /s/ occurred, measured as number of segments, was also analyzed. Terrell (1979) reports that /s/ in longer words in Cuban Spanish is more prone to phonetic reduction than /s/ in shorter words. He proposes that speakers may feel an unconscious need to maintain the articulation of shorter words as there is simply less phonetic material available and therefore misunderstandings may be more likely when there is reduction in shorter words than in longer ones.

The preceding and following phonological contexts were analyzed, where possible, as: high vowel, non-high vowel, coronal consonant, non-coronal consonant, pause. In the Barranquilla data, a preceding pause is not possible, as only word-medial, syllable-final /s/ and word-final /s/ is analyzed in that dialect. Further, while possible in word-final position, /s/ followed by a high vowel is not

attested in the Barranquilla data and, therefore, that level of the following segment is not represented in these data. Similarly, of the 1,749 tokens of /s/ in the Cali data, only 6 tokens are preceded by a non-coronal consonant and, as such, those tokens of /s/ were conflated with tokens preceded by coronal consonants. Several studies have shown that the preceding and following phonological contexts are two of the most influential factors in predicting the phonetic realization of /s/ in Spanish. Brown (2005) shows that the preceding context is the most influential factor in the prediction of syllable-initial /s/ realization while Brown (2009b) shows that the following context is most influential in the prediction of syllable- and word-final /s/ realization. Consequently, in a study such as this one, it is vitally important to statistically control for these two highly influential variables in order to accurately measure the relative influence of the two frequency measurements investigated here.

In addition to the surrounding phonological context, the syllabic stress of /s/ was taken into consideration. When word-final /s/ was followed by a vowel (with no intervening pause), resyllabification was assumed (e.g., *dos años* 'two years' > *do-sa-ños*) and stress was coded accordingly. Alba (1982) has shown that phonetic reduction is favored in unstressed syllables. In fact, the definition of stress itself cross-linguistically can safely be proposed to be the temporal elongation of a syllable accompanied by a full(er) articulation of the segments. Therefore, unstressed syllables display the opposite of this, that is, a shorter and less robust articulation of segments. This line of reasoning highlights the importance of controlling for this variable in the current study, which analyzes duration and robustness of articulation of /s/ as measured by the center of gravity of the sibilant sound and the amount of voicing.

Finally, the position of /s/ within the word and the syllable was analyzed. The word and the syllable positions were conflated as these two factors are not orthogonal. For example, word-initial /s/ is invariably in syllable-initial position while word-final /s/ can be in either syllable-initial or -final position, depending on the initial segment of the following word, whether a consonant, a vowel, or a pause. Due to their lack of orthogonality, these variables were conflated to avoid interaction between them in the multifactorial analyses.

In order to gauge the relative conditioning effect of the two frequency measurements in question, both monofactorial correlation tests as well as monofactorial and multifactorial linear regression analyses were run using the programming language and environment R (R Development Core Team 2013). Six multifactorial linear regressions are modeled, two models for each of the three dependent variables: for each dependent variable, one model included form frequency but excluded lemma frequency and the other model included lemma frequency but excluded form frequency. However, in all models the other predictor

variables shown in the literature to condition /s/ realization, as discussed above, were also included. As such, the only difference between the multifactorial linear regressions within a given dependent variable is the frequency measurement included.

Following Gries (2013: Ch. 5), a stepwise procedure is employed in the multifactorial linear regressions. Variables that do not significantly contribute to the prediction of the dependent variable in question are iteratively eliminated until only significant ones remain, giving the so-called *minimal adequate model* (cf. Gries 2013: 261). This procedure is employed for the multifactorial linear regressions reported below.

Given the hierarchical relationships that exist in the data (i.e., multiple observations from the same individual as well as multiple observations from the same word), the reader may find it curious that the decision was made to not employ mixed-effects models, as these models can account for hierarchical relationships. Mixed-effects models are based on the assumption that variances in the different groupings are equal. However, this assumption is not met by any of the groupings in these data (speaker or word), as evidenced by twelve Levene's tests for homogeneity of variance across the groupings.[4] In fact, given the structure of the sociolinguistic corpora used in this paper and sociolinguistic data based on spontaneous speech in general, some words occur only once in a sample and therefore have a variance of zero, while others occur more than once and therefore generally have a variance greater than zero.

With the coding of the three dependent variables and the eight predictor variables complete, the monofactorial correlation tests and monofactorial and multifactorial linear regression analyses were run, as reported in Section 3.

3 Results

The results of a series of monofactorial and multifactorial analyses demonstrate that form frequency has a stronger conditioning effect than lemma frequency in the realization of /s/ in the Cali data. In the Barranquilla data however, there is virtually no difference between the effect of the two frequency measurements, and in the end, neither measurement influences /s/ realization in this variety.

4 The p-values of the twelve Levene's tests of homogeneity of variance (two dialects by three dependent variables by two groupings) are all below 0.001, except for the test that measured /s/ duration in Cali grouped by speaker, which yields a p-value of 0.005.

3.1 Cali

The results of the monofactorial correlation tests and the multifactorial linear regression analyses of /s/ realization in the Cali data consistently suggest a stronger influence from form frequency than from lemma frequency in all three dependent variables.

3.1.1 Duration

The duration of /s/ is affected more by form frequency than by lemma frequency. Figure 2 displays the distribution of /s/ duration by each frequency measurement, with form frequency on the left side of the figure and lemma frequency on the right side.[5] The sloped dotted line in the scatterplots is a straight regression line based on a monofactorial linear regression model. The curved solid line is a locally-based lowess line. The vertical and horizontal tightly dotted lines mark the median frequency and /s/ duration, respectively, and divide each scatterplot into four quadrants. The number in the corner of each quadrant notes the number of data points within that quadrant.[6] The size of the numbers is proportional to the number of data points in each quadrant.

As seen in Figure 2, both form and lemma frequencies condition /s/ duration. The straight dotted regression line has a downward slope indicating a negative relation between /s/ duration and frequency. As expected, as frequency increases, /s/ duration decreases. In addition, the number of data points in the upper left and bottom right quadrants is larger than the number of data points in the other quadrants, which, again, indicates a negative relation between /s/ duration and frequency.

Upon comparison of the two frequency measurements, it is apparent that form frequency has a larger conditioning effect than lemma frequency. The slope of the regression line is steeper in the form frequency scatterplot than in that of lemma frequency; in two monofactorial linear regressions, one with form

5 Thanks are expressed to Stefan Th. Gries for the R code that this and subsequent scatterplots are based on.

6 Rather than through visual inspection, the numbers in the four quadrants were calculated by finding the number of tokens above and below the medians of the two variables. For example, the numbers in the top-left quadrants of Figure 1 were calculated by finding the number of tokens with a duration of /s/ above the median duration and a frequency below the median frequency.

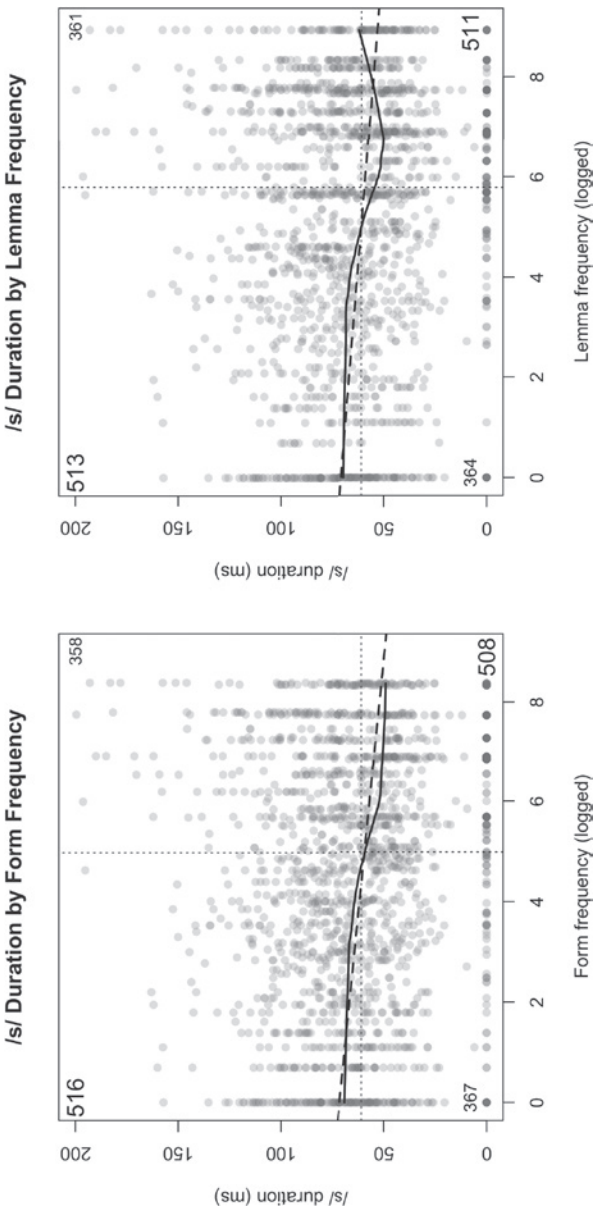

Fig. 2: Scatterplots of /s/ duration by frequency in Cali

frequency converted to a z-standardized natural logarithm number and the other regression with lemma frequency also transformed, the slope of form frequency is steeper (coefficient = −0.0062, $R^2 = 0.028$, $p \leq 0.001$) than that of lemma frequency (coefficient = −0.0051, $R^2 = 0.019$, $p \leq 0.001$). The lowess line in the form frequency scatterplot is more consistently negatively sloped.

In addition to the information gleaned from Figure 2, Kendall's rank correlation tau tests provide further evidence that form frequency has a larger conditioning effect than lemma frequency. Form frequency has a larger absolute tau value (tau = −0.125; $p \leq 0.001$) than lemma frequency (tau = −0.091; $p \leq 0.001$). Kendall's rank correlation tau test was chosen as the assumption of normality among the data that is required to run a Pearson's product-moment correlation test was not met, as evidenced by two Shapiro-Wilk normality tests (form frequency: $W = 0.66$, $p \leq 0.001$; lemma frequency: $W = 0.66$, $p \leq 0.001$). For this reason, Kendall's rank correlation tau tests were also run for the other dependent variables discussed below, in both the Cali data as well as the Barranquilla data.

In addition to the monofactorial correlation analyses, two multifactorial linear regressions were run, one that included form frequency but excluded lemma frequency, and the other that included lemma frequency but excluded form frequency. The results of these two linear regressions suggest that, as shown in the monofactorial analyses, form frequency is more influential than lemma frequency. In the model that included form frequency but excluded lemma frequency, form frequency, along with other predictor variables, was selected as a making a significant contribution to /s/ duration (model statistics: $R^2 = 0.2481$, $p \leq 0.001$). On the other hand, in the model that included lemma frequency but excluded form frequency, lemma frequency was not selected as making a significant contribution to /s/ duration (model statistics: $R^2 = 0.2478$, $p \leq 0.001$).

It should be noted that, as seen in Figure 2, there is an upswing in the locally-based lowess line among the highest-frequency lemmas in the lemma frequency scatterplot. This is caused by low-frequency forms belonging to high-frequency lemmas. For example, the low-frequency form *estaban* 'be imperfect indicative 3p' belongs to the high-frequency lemma *ESTAR* 'to be'. The generally longer duration of low-frequency forms causes the high-frequency lemmas to have a longer average duration. This trend is also seen below in the other two dependent variables in these Cali data.

In summary, the results from both monofactorial correlation tests as well as the monofactorial and multifactorial linear regressions suggest that form frequency has a larger conditioning effect than lemma frequency on /s/ duration in these Cali data.

3.1.2 Center of gravity

The results of monofactorial correlation and linear regression analyses as well as multifactorial linear regression analyses of the center of gravity of /s/ in these Cali data are similar to those seen with /s/ duration: form frequency has a larger conditioning effect than lemma frequency. Figure 3 below presents a portrait of the center of gravity of /s/ that is similar to /s/ duration in Figure 2 above. The lines and numbers in Figure 3 are the same as those in Figure 2, described above.

Both form and lemma frequencies condition the center of gravity of /s/. The negatively sloped straight regression line indicates a negative relation between center of gravity and frequency in both scatterplots. Likewise, the upper left and bottom right quadrants of both figures contain more data points than the other quadrants, again, indicative of a negative correlation between center of gravity and frequency, whether form frequency or lemma frequency.

When compared to each other, form frequency and lemma frequency show differences in their conditioning effect on the center of gravity of /s/. The slope of the regression line is steeper in the form frequency scatterplot than in that of lemma frequency; in two monofactorial linear regressions, one with form frequency converted to a z-standardized natural logarithm number and the other regression with lemma frequency also transformed, the slope of form frequency is steeper (coefficient = –383.62, $R^2 = 0.0339$, $p \leq 0.001$) than that of lemma frequency (coefficient = –247.94, $R^2 = 0.0138$, $p \leq 0.001$). The lowess smoother line in the form frequency scatterplot is more consistently negatively sloped. Likewise, the sum of data points in the upper left and bottom right quadrants of the form frequency scatterplot is greater than in the lemma frequency scatterplot. Additionally, Kendall's rank correlation tests show a stronger correlation between center of gravity of /s/ and form frequency (tau = –0.108; $p \leq 0.001$) than with lemma frequency (tau = –0.049; $p = 0.003$).

In addition to the monofactorial analyses, two multifactorial linear regressions were run to measure the influence of each frequency measurement alongside the influence of the other predictor variables mentioned above. Specifically, one multifactorial linear regression includes form frequency, along with the other predictor variables, but excludes lemma frequency. A second model includes lemma frequency, along with other predictors, but excludes form frequency. In both cases, the frequency measurement in question was converted to a z-standardized natural logarithm number in order to directly compare the coefficients of the frequency measurements. Even though both form and lemma frequencies were selected as significant variables in their respective models, the results suggest a stronger conditioning effect from form frequency (form frequency

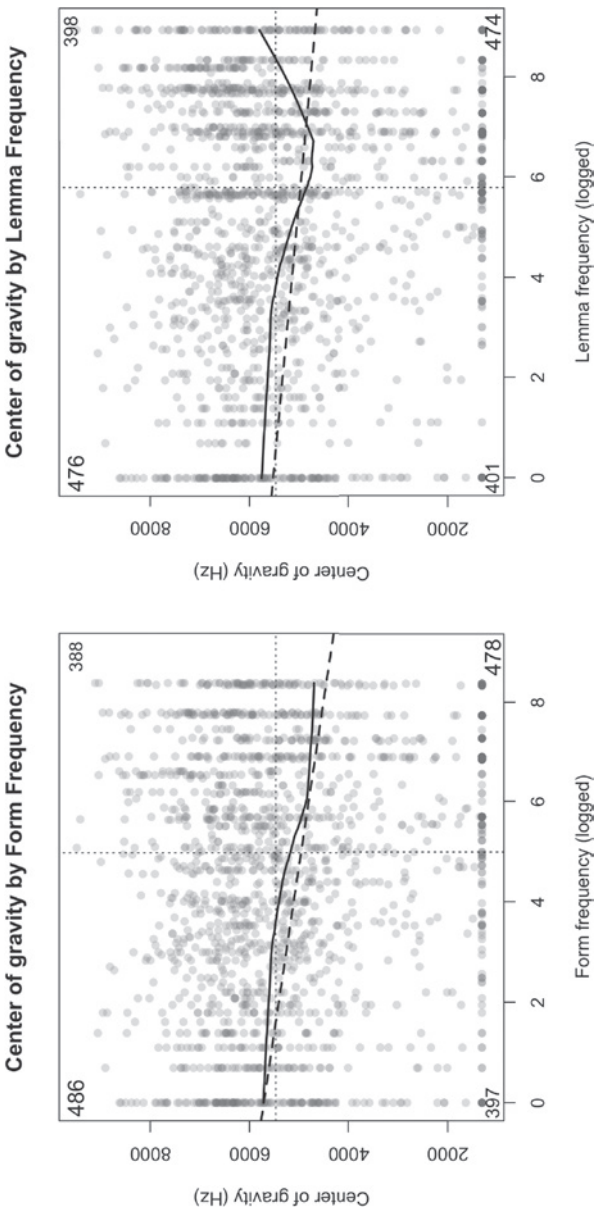

Fig. 3: Scatterplots of center of gravity of /s/ by frequency in Cali

statistics: coefficient = –244.804, standard error = 63.304, t = –3.867, $p \leq 0.001$; model statistics: Adjusted $R^2 = 0.189$, $F_{14,\ 1,734} = 30.23$, $p \leq 0.001$) than from lemma frequency (lemma frequency statistics: coefficient = –146.01, standard error = 59.73, t = –2.445, p = 0.015; model statistics: Adjusted $R^2 = 0.185$, $F_{15,\ 1,733} = 27.46$, $p \leq 0.001$) with the center of gravity of /s/ in these Cali data.

In summary, the results from both monofactorial correlation tests as well as the monofactorial and multifactorial linear regressions suggest that form frequency has a larger conditioning effect than lemma frequency on the center of gravity of /s/ in these Cali data.

3.1.3 Voicelessness

Similar to the other two dependent variables, the voicelessness of /s/ is more influenced by form frequency than by lemma frequency (see Figure 4).

As with the previous two dependent variables, the straight regression line in the form frequency scatterplot is steeper than the line in the lemma frequency scatterplot; in two monofactorial linear regressions, one with form frequency converted to a z-standardized natural logarithm number and the other regression with lemma frequency converted to a z-standardized natural logarithm number, the slope of form frequency is steeper (coefficient = –0.0417, $R^2 = 0.0157$, $p \leq 0.001$) than that of lemma frequency (coefficient = –0.0275, $R^2 = 0.0065$, $p \leq 0.001$). As expected, the sum of the data points in the upper left and lower right quadrants is greater in the form frequency scatterplot than in the lemma frequency one. Similarly, the curved locally-based lowess line is more consistently negatively sloped in the form frequency scatterplot than in the lemma frequency one. Further, the absolute value of the Kendall's rank correlation tau test is larger for form frequency (tau = –0.102; $p \leq 0.001$) than for lemma frequency (tau = –0.058; $p \leq 0.001$).

Likewise, two multifactorial linear regressions show a stronger effect from form frequency than from lemma frequency. In the run that includes form frequency but excludes lemma frequency, form frequency, along with other predictor variables, is selected as a making a significant contribution to /s/ duration (model statistics: $R^2 = 0.159$, $p \leq 0.001$). On the other hand, in the model that includes lemma frequency but excludes form frequency, lemma frequency is not selected as making a significant contribution to /s/ duration (model statistics: $R^2 = 0.156$, $p \leq 0.001$).

In summary, in these Cali data, it is clear that there is a stronger conditioning effect from form frequency than from lemma frequency on the realization of /s/, whether it be duration, center of gravity, or the percentage of voicelessness.

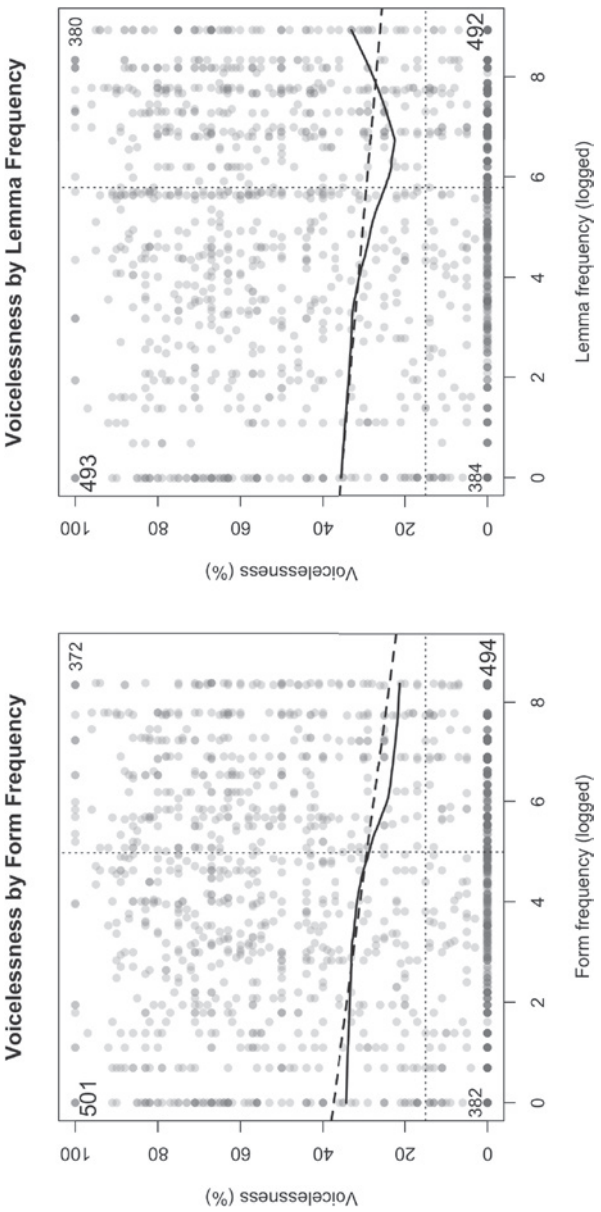

Fig. 4: Scatterplots of voicelessness of /s/ by frequency in Cali

3.2 Barranquilla

Unlike in the Cali data, there is no difference in the conditioning effect of form frequency versus lemma frequency in the Barranquilla data, as neither frequency type has a significant effect on /s/ realization.

3.2.1 Duration

As seen in Figure 5, /s/ duration in Barranquilla is basically unaffected by either frequency measurement. The straight regression line as well as the locally-based lowess line of both scatterplots are virtually flat; two monofactorial linear regressions with the frequency measurements transformed to z-standardized natural logarithm numbers show no effect from either form frequency (coefficient = 0.003, $R^2 = -0.00092$, $p = 0.758$) or lemma frequency (coefficient = 0.006, $R^2 = -0.0006$, $p = 0.53$). The number of data points in each of the four quadrants in both scatterplots is nearly identical. Further, Kendall's tau tests show no significant correlation between /s/ duration and either form frequency (tau = −0.03, $p = 0.2$) or lemma frequency (tau = −0.02, $p = 0.4$). Additionally, when run separately, neither frequency measurement is selected as significant in multifactorial linear regressions with other predictor variables (form frequency model: $R^2 = -0.34$, $p \leq 0.001$; lemma frequency model: $R^2 = -0.34$, $p \leq 0.001$). In summary, frequency has no effect on /s/ duration in these Barranquilla data.

3.2.2 Center of gravity

The center of gravity of /s/ in Barranquilla is also not affected by either frequency measurement (see Figure 6). The regression lines are basically flat; two monofactorial linear regressions fail to return a significantly inclined slope for either frequency measurement (form frequency model: $R^2 = -0.000003$, $p = 0.3$; lemma frequency model: $R^2 = -0.0007$, $p = 0.6$). The lowess lines are mostly flat, aside from highest-frequency tokens. While the number of data points in the upper left and bottom right quadrant is larger than those in other two quadrants, Kendall's tau tests fail to return a significant effect from frequency: tau value of form frequency test = −0.03 ($p = 0.2$), tau value of lemma frequency test = −0.02 ($p = 0.4$). Additionally, when run separately, neither frequency measurement is selected as significant in multifactorial linear regressions with other predictor variables (form frequency model: $R^2 = 0.21$, $p \leq 0.001$; lemma frequency model: $R^2 = 0.34$, $p \leq 0.001$). To summarize, frequency has no effect on the center of gravity of /s/ in these Barranquilla data.

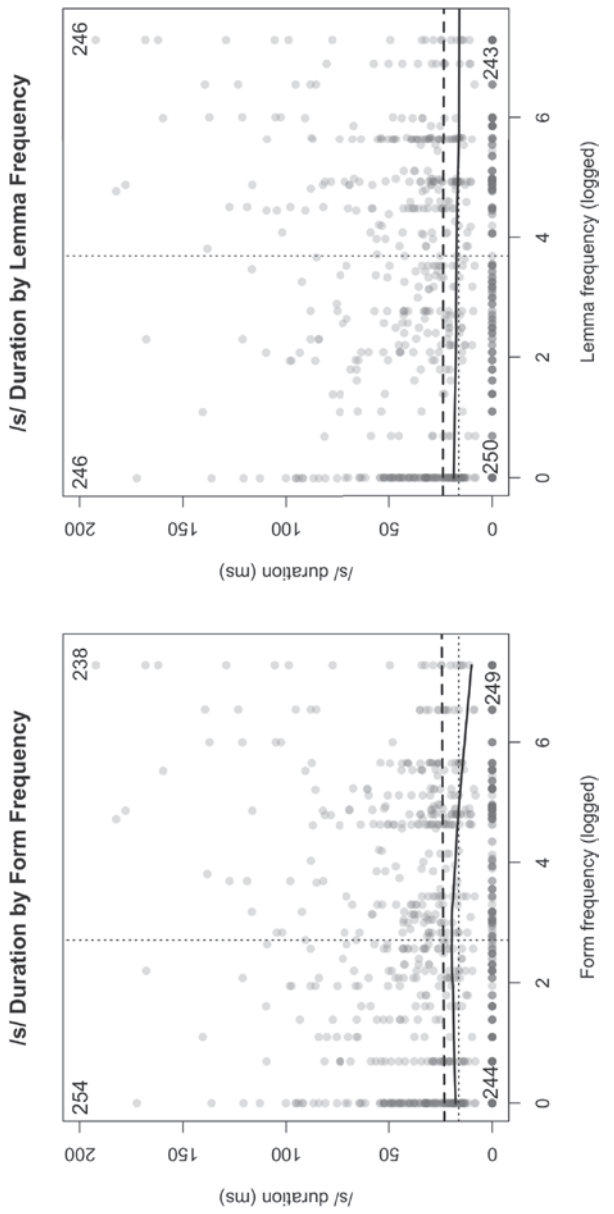

Fig. 5: Scatterplots of /s/ duration by frequency in Barranquilla

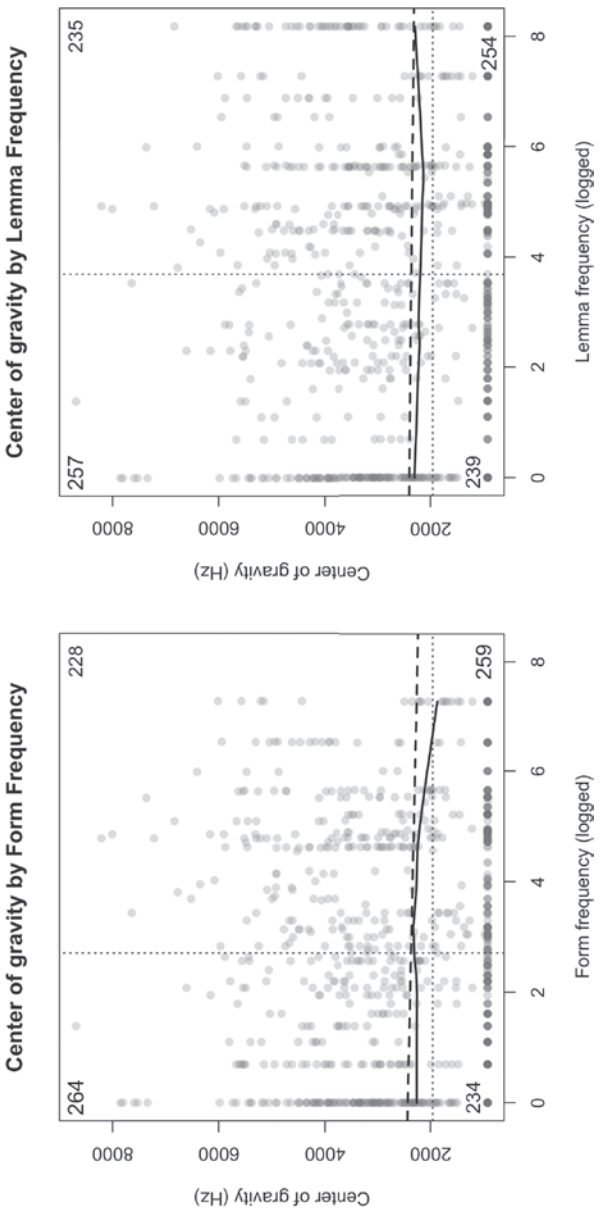

Fig. 6: Scatterplots of center of gravity of /s/ in Barranquilla

3.2.3 Voicelessness

Like the previous two dependent variables, the percentage of voicelessness of /s/ in these Barranquilla data is not affected by either frequency measurement (see Figure 7).

Again, the straight regression lines are not significantly sloped; in two monofactorial linear regressions, neither form frequency ($R^2 = -0.0009$, $p = 0.8$) nor lemma frequency ($R^2 = -0.0006$, $p = 0.5$) significantly affect voicelessness of /s/. Even though (unexpectedly) there are fewer data points in the upper left and lower right quadrants, Kendall's rank correlation tau tests do not show a significant effect from either form frequency (tau = 0.002, $p = 0.9$) or lemma frequency (tau = 0.009, $p = 0.7$). Additionally, when run separately, neither frequency measurement is selected as significant in multifactorial linear regressions with other predictor variables (form frequency model: $R^2 = 0.1$, $p \leq 0.001$; lemma frequency model: $R^2 = 0.1$, $p \leq 0.001$). To summarize, frequency has no effect on the percentage of voicelessness of /s/ in these Barranquilla data.

In the end, in these Barranquilla data, the frequency measurements fail to display a significant conditioning effect on the realization of /s/, regardless of whether it is measured by duration, center of gravity, or percentage of voicelessness.

4 Discussion

Two findings stand out in the Cali and Barranquilla data analyzed in this paper: (1) frequency plays a significant role in the prediction of /s/ realization in the Cali data, but not in the Barranquilla data, and (2) among the tokens in the Cali data, form frequency has a stronger conditioning effect than lemma frequency. As such, in response to Research Question 1, this paper provides evidence that form frequency has a larger conditioning effect than lemma frequency. In theoretical terms, to the degree that form frequency is a proxy for the activation state of a form in memory while lemma frequency is a proxy for the combined activation states of members of a paradigm (cf. Pierrehumbert 2001; Vannest et al. 2011), the results of this paper suggest that the activation state or strength of a single word in the lexicon is more influential than the combined strength of all members of the paradigm to which a word belongs.

On the other hand, in response to Research Question 2, this paper does not provide evidence that frequency has the same effect in the two varieties studied. Rather, the results show that the frequency measurements have different effects on the phonetic realization of /s/ in the two varieties of Colombian Spanish.

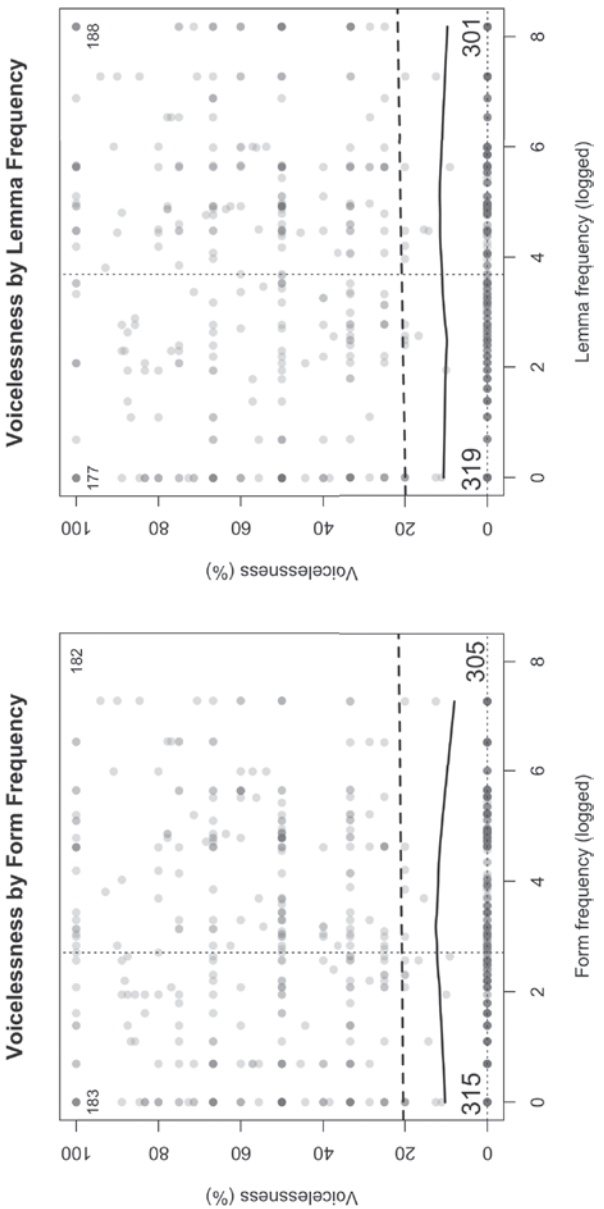

Fig. 7: Scatterplots of the voicelessness of /s/ by frequency in Barranquilla

4.1 Cali versus Barranquilla

Why is /s/ in the speech of Cali affected by frequency while /s/ in the speech of Barranquilla is not? Bybee (2000: 81) proposes that "... sound change, or phonetic change, affects the lexicon gradually by spreading out across lexical items from the most frequent to the least frequent...." One possible explanation for the curvature of the locally-based lowess line in the form frequency scatterplots in the Cali data (cf. Figures 2, 3, and 4 above) may be as follows: The highest-frequency words have already been affected by this sound change[7] while the lowest-frequency words have not yet been affected or have been affected to a lesser degree. As such, if this proposition is true, it appears that this sound change is making its way through the middle-frequency words in these Cali data. Differently, it is proposed that this sound change has already affected the speech of Barranquilla to such a degree that frequency cannot exert any more influence.

As noted in Section 2, Cali displays a low rate of overall /s/ reduction while Barranquilla displays a high rate of reduction. Barranquilla is located on the northern coast of Colombia and is generally considered to be a Caribbean or Caribbean-like dialect of Spanish. Cali, on the other hand, is located inland in south-central Colombia and can likely be categorized as a highland dialect. The fact that /s/ in Barranquilla, a high-reducing dialect, is not affected by frequency concurs with a growing literature showing that frequency effects are less likely to be observed when a sound change is nearing completion (cf. Labov 1994: 65; Phillips 2006: 13; Brown 2009b: 207). Specifically concerning Spanish, Brown (2009b) finds that in two dialects with high levels of /s/ reduction (San Juan, Puerto Rico and Mérida, Venezuela), frequency is either not influential or minimally influential. In the same study, Brown shows that /s/ in two dialects with lower levels of overall /s/ reduction (northern New Mexico and southern Colorado colonial Spanish and Cali, Colombia[8]) is significantly affected by frequency. It should be noted that all predictor factors, not just frequency, will likely lose conditioning effect as a linguistic change nears completion.[9]

Although the results of this paper concur with previous literature, and specifically Brown (2009b) with regards to Spanish, the finding that frequency plays no role in /s/ weakening in the Barranquilla data contradicts the results of File-Muriel (2009), a study that also analyzes /s/ in Barranquilla Spanish. However, several factors distinguish that study from the present one. First, File-Muriel

7 The question of whether the variable s-realization reported here represents a sound change in progress or stable variation is out of the scope of this paper and should be taken up elsewhere.
8 These results were found using a different corpus of Cali, Colombia (cf. Travis 2005).
9 Thanks are expressed to an anonymous reviewer for pointing this out.

(2009) analyzes only word-medial, syllable-final /s/ while this study analyzes both word-medial, syllable-final /s/ and word-final /s/. In addition to syllable-final position, word-final /s/ can be in syllable-initial position when followed by a vowel, as resyllabification is the norm in Spanish. Second, since File-Muriel (2009) analyzes only word-medial /s/, for example *hasta* 'until', all tokens in his study are examples of lexical /s/. In contrast, the present study analyzes both lexical /s/ and morphological /s/, whether plural /s/ (*casas* 'houses') or verbal /s/ (*hablamos* '1p present talk', *hablas* '2s present talk'). Third, File-Muriel (2009) and the present study analyze different registers of speech. While File-Muriel utilizes a reading task of words extracted from the Colombian newspaper *El Tiempo*, the present study measures /s/ in spontaneous speech in sociolinguistic interviews. Alba (2004) demonstrates that /s/ realization covaries with differences in register. An epiphenomenon of this difference in register is that the present study analyzes more high-frequency words than does File-Muriel (2009). In general, spontaneous speech is rife with high-frequency words while other, less spontaneous, speech is not. Further, File-Muriel's methodology specifically elicits the production of more low-frequency words than would otherwise be the case in order to more explicitly compare low- and high-frequency words. These differences seem to allow frequency to exert a significant effect on /s/ realization in File-Muriel's study but then lack a significant conditioning effect in the present study. Additionally, File-Muriel's study employs auditory acoustic analysis, not instrumental acoustic measurements, making it vulnerable to a host of issues discussed in previous literature (cf. Erker 2010; File-Muriel and Brown 2011).

4.2 Form versus lemma frequency

The second major finding of this paper is that form frequency is more influential than lemma frequency on the phonetic realization of /s/ in the Cali data. This suggests that the frequency information stored in the exemplar cluster of an individual word has a larger conditioning effect than the frequency information stored with other members in an inflectional paradigm, as transmitted through associative networks (Bybee 1985, 2010). Consequently, even though there may exist an effect exerted by the frequency of other words in a paradigm, that is by lemma frequency, the influence of the frequency information stored with individual words, that is form frequency, appears to be more important. To investigate this possibility further, three additional linear regressions were modeled in the Cali data, one for each of the three dependent variables, in which both form frequency and lemma frequency were included in the same model. Interestingly, in two of the three additional linear regressions, specifically in the model of the

center of gravity of /s/ and in the model of the percentage of voicelessness of /s/, form frequency was selected as making a significant contribution to the dependent variable in question while lemma frequency was not (center of gravity of /s/ model: $R^2 = 0.189$, $p \leq 0.001$; percentage of voicelessness of /s/ model: $R^2 = 0.159$, $p \leq 0.001$). In the third model, that of the duration of /s/, neither frequency measurement was selected as significant when both were included.[10]

Inasmuch as frequency is a proxy for the resting activation state of words in memory (cf. Pierrehumbert 2001; Vannest et al. 2011), the results of this study suggest that the level of activation of individual word forms is more predictive of the phonetic realization of /s/ than the activation states of the other members of a lemma. Therefore, although the combined activation states of other members of a paradigm may influence this particular linguistic variable, the activation state of the individual forms is likely more influential.

5 Conclusion

This paper has measured and explicitly compared the relative conditioning effects of form and lemma frequencies on /s/ realization in two varieties of Colombian Spanish. Two main results are highlighted: (1) frequency has a conditioning effect on /s/ in the speech of Cali but not in the speech of Barranquilla, and (2) form frequency exerts a greater influence than lemma frequency. The second of the two main results suggests that the influence of the individual forms on /s/ realization is greater than the influence of the combined frequencies of members of the same paradigm. In short, while there may be an influence from other members of a paradigm, in spontaneous speech the frequency of the individual word forms is more predictive of /s/ realization. In general terms, the results of this paper suggest that the frequency information stored on individual forms of the lexicon is more important during a sound change than is the influence from sister members of a paradigm.

References

Alba, Matthew C. 2006. Accounting for Variability in the Production of Spanish Vowel Sequences. In Nuria Sagarra & Almeida Jacqueline Toribio (eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, 273–285. Somerville, MA: Cascadilla Proceedings Project.

10 This despite the fact that form frequency was nearly selected as a significant variable with a p-value of 0.058.

Alba, Orlando. 1982. Función del acento en el proceso de elisión de la /s/ en la República Dominicana. *El español del Caribe*, 15–26. Santiago, Dominican Republic: Universidad Católica Madre y Maestra.

Alba, Orlando. 2004. *Cómo Hablamos Los Dominicanos: Un Enfoque Sociolingüístico*. (Colección Centenario). Santo Domingo, Dominican Republic: Grupo León Jimenes.

Bayley, Robert, Kristen Ware & Cory Holland. 2013. Lexical Frequency and Syntactic Variation: A Test of a Linguistic Hypothesis. *University of Pennsylvania Working Papers in Linguistics* 19(2). http://repository.upenn.edu/pwpl/vol19/iss2/.

Boersma, Paul & David Weenink. 2012. *Praat: doing phonetics by computer*. www.praat.org.

Brown, Earl K. 2009a. The Relative Importance of Lexical Frequency in Syllable- and Word-Final /s/ Reduction in Cali, Colombia. In Joseph Collentine, Maryellen García, Barbara Lafford & Francisco Marcos Marín (eds.), *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, 165–178. Somerville, MA: Cascadilla Proceedings Project. http://www.lingref.com/cpp/hls/11/index.html.

Brown, Earl K. 2009b. *A Usage-Based Account of Syllable- and Word-Final /s/ Reduction in Four Dialects of Spanish*. LINCOM Studies in Romance Linguistics 62. Munich: Lincom Europa. http://www.lincom.eu.

Brown, Esther L. 2005. New Mexican Spanish: Insight into the Variable Reduction of "la ehe inihial" (/s-/). *Hispania* 88(4). 813–824.

Brown, Esther L. & Rena Torres Cacoullos. 2003. Spanish /s/: A different story from beginning (initial) to end (final). *A Romance Perspective in Language Knowledge and Use. Selected Papers from the 31th Linguistic Symposium on Romance Languages (LSRL), Chicago, 19–22 April 2001 (Current Issues in Linguistic Theory (CILT) 238)*, 22–38. Amsterdam/ Philadelphia: John Benjamins.

Bush, Nathan. 2001. Frequency effects and word-boundary palatalization in English. In Joan Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 255–280. Amsterdam: John Benjamins.

Bybee, Joan. 1985. *Morphology: A Study of the Relation Between Meaning and Form*. (Typological Studies in Language, V. 9). Philadelphia: John Benjamins.

Bybee, Joan. 2000. The phonology of the lexicon: evidence from lexical diffusion. *Usage-based models of language*, 65–85. Stanford, CA: Center for the Study of Language and Information.

Bybee, Joan. 2006. From Usage to Grammar: The Mind's Response to Repetition. *Language* 82(4). 711–733.

Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge University Press.

Bybee, Joan & Joanne Scheibman. 1999. The Effect of Usage on Degree of Constituency: The Reduction of don't in American English. *Linguistics* 37. 575–596.

Cresti, Emanuela & Massimo Moneglia. 2005. *C-ORAL-ROM: Integrated reference corpora for spoken Romance languages*. Amsterdam: John Benjamins.

Diaz-Campos, Manuel, Stephen Fafulas & Michael Gradoville. 2012. Variable Degrees of Constituency: Frequency Effects in the Alternation of pa vs. para in Spoken Discourse. In Kimberly L. Geeslin & Manuel Diaz-Campos (eds.), *Selected Proceedings of the 14th Hispanic Linguistics Symposium*, 75–87. Somerville, MA: Cascadilla Proceedings Project.

Diaz-Campos, Manuel & Michael Gradoville. 2011. An Analysis of Frequency as a Factor Contributing to the Diffusion of Variable Phenomena: Evidence from Spanish Data. In Luis Ortiz-López (ed.), *Selected Proceedings of the 13th Hispanic Linguistics Symposium*, 224–238. Somerville, MA: Cascadilla Proceedings Project.

Dinkin, Aaron J. 2008. The real effect of word frequency on phonetic variation. *University of Pennsylvania Working Papers in Linguistics* 14(1). http://repository.upenn.edu/pwpl/vol14/iss1/.

Eddington, David. 2011. What are the contextual phonetic variants of /β, ð, ɣ/ in colloquial Spanish? *Probus* 23(1). 1–19.

Erker, Daniel. 2010. A subsegmental approach to coda /s/ weakening in Dominican Spanish. *The International Journal of the Sociology of Language* 203. 9–26.

Erker, Daniel & Gregory R. Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3). 526–557.

File-Muriel, Richard. 2009. The role of lexical frequency in the weakening of syllable-final lexical /s/ in the Spanish of Barranquilla, Colombia. *Hispania* 92(2). 348–360.

File-Muriel, Richard & Earl K. Brown. 2010. The gradient nature of /s/-lenition in Caleño Spanish. *University of Pennsylvania Working Papers in Linguistics* 16(2). 46–55.

File-Muriel, Richard & Earl K. Brown. 2011. The Gradient Nature of s-Lenition in Caleño Spanish. *Language Variation and Change* 23(2). 223–243.

Gahl, Susanne. 2008. *Time* and *Thyme* Are Not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech. *Language: Journal of the Linguistic Society of America* 84(3). 474–496.

Gradoville, Michael Stephen. 2011. Validity in Measurements of Fricative Voicing: Evidence from Argentine Spanish. In Scott M. Alvord (ed.), *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology*, 59–74. Somerville, MA: Cascadilla Proceedings Project.

Gries, Stefan Th. 2013. *Statistics for Linguistics with R: A Practical Introduction*. 2nd edn. Berlin: De Gruyter Mouton.

Hooper, Joan Bybee. 1976. Word frequency in lexical diffusion and the source of morphophonological change. *Current Progress in Historical Linguistics*, 96–105. Amsterdam: North Holland.

Jescheniak, Jörg D. & Willem J. M. Levelt. 1994. Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20(4). 824–843.

Jurafsky, Daniel, Alan Bell & Cynthia Girand. 2002. The Role of the Lemma in Form Variation. In Carlos Gussenhoven & Natasha Warner (eds.), *Papers in Laboratory Phonology VII*, 1–34. Berlin/New York: Mouton de Gruyter.

Labov, William. 1984. Field Methods of the Project on Linguistic Change and Variation. In J. Baugh & J. Sherzer (eds.), *Language in use: Readings in Sociolinguistics*, 28–53. Englewood Cliffs, NJ: Prentice Hall.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

Lipski, John. 1994. *Latin American Spanish*. New York: Longman.

Minnick Fox, Michelle Annette. 2006. Usage-based Effects in Latin American Spanish Syllable-final /s/ Lenition. Philadelphia: University of Pennsylvania.

Nosofsky, Robert M. 1988. Similarity, frequency, and category representations. *Journal of Experimental Psychology: learning, memory, and cognition* 14. 54–65.

Phillips, Betty. 1984. Word frequency and the actuation of sound change. *Language* 60. 320–342.

Phillips, Betty. 2006. *Word frequency and lexical diffusion*. Houndsmill and New York: Palgrave MacMillan.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 137–157. Amsterdam/Philadelphia: John Benjamins Publishing Company.
R Development Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.
Scheibman, Joanne. 2000. I dunno: A Usage-Based Account of the Phonological Reduction of don't in American English Conversation. *Journal of Pragmatics* 32(1). 105–124.
Schuchardt, H. 1885. *Über die Lautgesetze: gegen die Junggrammatiker*. Berlin: Oppenheim.
Schwegler, Armin, Juergen Kempff & Ana Ameal-Guerra. 2010. *Fonética y fonología españolas*. 4th ed. New York: John Wiley and Sons.
Terrell, Tracy D. 1979. Final /s/ in Cuban Spanish. *Hispania* 62. 599–612.
Travis, Catherine E. 2005. *Discourse Markers In Colombian Spanish: A Study In Polysemy*. (Cognitive Linguistics Research). Berlin: Mouton de Gruyter.
Vannest, Jennifer, Elissa L. Newport, Aaron J. Newman & Daphne Bavelier. 2011. Interplay between morphology and frequency in lexical access: The case of the base frequency effect. *Brain Research* 1373. 144–159.
Vennemann, Theo & Terence H. Wilbur (eds.). 1972. *Schuchardt, the Neogrammarians, and the Transformational Theory of phonological change: Four essays*. Frankfort: Athenaum.
Walker, James A. 2012. Form, function, and frequency in phonological variation. *Language Variation and Change* 24(3). 397–415.

Bionotes

Earl K. Brown received a PhD in Hispanic Linguistics in 2008 at the University of New Mexico, Albuquerque. He currently works as an Assistant Professor of Spanish Linguistics at Kansas State University, Manhattan. His research centers on language variation in Spanish, especially its sound system, and often utilizes corpus linguistics techniques.

Michael S. Gradoville finished a dual PhD in both Hispanic Linguistics and Linguistics (general) in 2013 at Indiana University, Bloomington. He currently works as an Instructor of Spanish at Spelman College in Atlanta, Georgia. His research focuses on language variation, linguistic theory, and research methodology based on data in Spanish and Portuguese.

Richard J. File-Muriel received a PhD in Hispanic Linguistics in 2007 at Indiana University, Bloomington. He currently works as an Assistant Professor of Hispanic Linguistics at the University of New Mexico, Albuquerque. His research interests are language variation and change, sociophonetic variation, and phonological theory.