

Paradigmatic Peer-Pressure: Word-Medial, Syllable-Initial /s/ Lenition in Dominican Spanish

Earl K. Brown

California State University, Monterey Bay

1. Introduction

Usage-based phonology (Bybee 2001, 2006, 2010), which is based on exemplar theory (Pierrehumbert 2001, 2003), proposes a memory capacity so expansive that, in theory, all tokens experienced in life could be mapped onto their respective exemplar clouds. However, if a token is similar enough to an existing one, it is simply mapped directly onto that existing token, reinforcing it in memory. This model proposes a very limited need for generative rules to produce the surface form, as the surface forms themselves are stored whole in memory and therefore can be directly accessed during production. It follows that derived words need not be formed by means of a base form, especially if the derived form is of a high enough frequency or if the derived form has a higher frequency than its base form (Hay 2001). Similarly, inflected words within a paradigm may gain autonomy if they are of a high enough frequency (Bybee 1985).

Further, Bybee (1985, 2001, 2010) proposes networks of associations between phonetically and semantically related words, such as members of the same inflectional paradigm and derivationally related words. It is proposed that high-frequency words can be part of the access route of retrieval during the production of related low-frequency words. This process can lead to the regularization of these low- or lower-frequency words. Hence, as used in this paper, "paradigmatic peer-pressure" refers to the influence that high- or higher-frequency words can exert over related low- or lower-frequency words, whether they are related by inflection or derivation. The access route of retrieval refers to which of the two routes a speaker (subconsciously) takes to access inflected or derived words: the compositional route or the whole-form route. The compositional route refers to an inflected or derived word being accessed by means of its component morphemes. The whole-form route allows complex words to be accessed directly, bypassing the component parts.

Examples of this paradigmatic peer-pressure given by Bybee (2001) include the English pairs *house-houses* and *louse-lice*. Because *house* is more frequent than *houses* (187,654 hits to 25,175, or seven times more frequent in the Corpus of Contemporary American English [COCA], Davies 2009), it can be part of the access route of the plural, and can therefore lead to the regularization of the etymologically voiced sibilant to a voiceless one: *hou[z]es* to *hou[s]es*. Similarly, because the plural *lice* is more frequent than the singular *louse* (606 hits to 133, or four times more frequent in the same corpus) *lice* is part of the access route of *louse* and has, in fact, replaced the etymological diphthong in *louse* with the one in *lice* ([laʊs] to [ləɪs]), as the average speaker is unaware that *louse* is (or more appropriately, was) the singular of *lice*. Contrary to these examples in which a more-frequent member of a paradigm affects a less-frequent member, the words in the pair *child-children* show no change in their respective root vowels; both maintain their different base vowels: ch[aɪ]ld and ch[I]ldren. There are two likely reasons that they retain their vowels. First, the difference in frequencies of the two words is not large enough for one to form part of the access route of the other (*children* is not quite twice as frequent as *child* in the COCA: 231,314 hits to 118,601). Secondly, the overall frequency of the singular may allow it to gain autonomy within the paradigm and thus resist morpho-phonemic

* I thank the faculty of the Department of Spanish and Portuguese at Brigham Young University for allowing me to use the Dominican data. Their generosity is greatly appreciated.

change (see Bybee 1985:57).

In addition to inflectional morphology, derivational morphology provides further examples of the effects of frequency differences between related words. Hay (2001) demonstrates with experimental data that higher-frequency derived words can be accessed whole during production rather than by way of their component parts, that is, the base and the affix morphemes. For example, the derived form *insane* is more frequent than its base form *sane*. As such, speakers need not access the prefix *in-* and the base *-sane* separately, but can directly access the morphologically complex derived form *insane*. Similarly, more-frequent derived words are less easily decomposed by speakers than their less-frequent derivationally related counterparts. Hay argues that "importantly, because words compete, the absolute frequency of the derived form is not so important as its frequency relative to the base form with which it is competing" (p. 1045). Figure 1 displays a visual representation of this proposal. The thicker circle of the derived form *insane* indicates a higher resting activation state, caused by its higher frequency relative to *sane*, and the dashed line indicates the preferred access route.

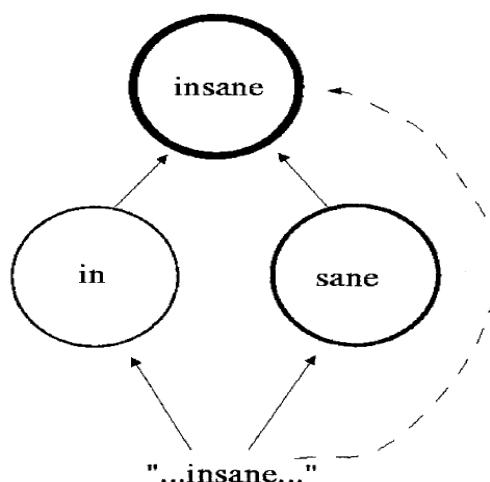


Figure 1: Whole-form access of a more-frequent derived word (from Hay 2001:1045)

Concerning this dual-route model, several researchers, including Hay (2001), Bybee (1985, 2010) and Baayen, Dijkstra, and Schreuder (1997), propose that whole-form access is likely with high-frequency derived forms whereas other researchers, including Pinker and Prince (1991) propose that this whole-form route is only chosen when the derived form is morphologically irregular. Regardless of how much influence morphological regularity has on which route is taken, most (functionalist) researchers agree that whole-form access is available, but that the compositional route is also available and, in fact, is likely with low-frequency or novel forms.

1.1. Lenition

The term lenition is used to describe the weakening process by which consonants become "less strongly occluded ... and/or more sonorous in some sense" (Lewis 2001:3). The motivation of this weakening process is usually ease of expression, or more technically, less articulatory effort or smaller articulatory gestures (Browman and Goldstein 1992). Simply put, that which is easier for the mouth to articulate will likely be articulated.

One way in which "weakened" sounds are changed is by receiving a shorter duration than would otherwise be the case (Lewis 2001:3). Intuitively, less articulatory effort should be required when less time is devoted to a given sound.

Another manifestation of this ease of articulation is assimilation. Pharies (2007) points out that a major step in the evolution of Spanish from Latin was the lenition of intervocalic voiceless stops to voiced ones, as seen in, for example, Latin *lupus* 'wolf' and *patrem* 'father' to Spanish *l**o**bo* and *pad**r**e*.

The voiceless consonants assimilated to the voicing of the surrounding vowels. (Synchronic examples of this same phenomenon in modern-day Spanish are detailed in Lewis (2001), and include *pecadores* → *pe[g]adores* 'sinners', taken from Dykstra (1955), and *melocotonero* → *melo[g]otonero* 'peach tree', taken from Salvador (1968).)

Another way in which less articulatory effort is manifested is in a more open articulation of consonants, which results in a more sonorous pronunciation. Spirantization, or the process that, generally speaking, converts stops to fricatives, also played a role in the development of Spanish from Latin and is seen in the examples presented above. The newly voiced stops in *padre* and *lobo* (and many other words with intervocalic voiced stops) became fricatives. That is, the articulation of these consonants became more open.

In summary, lenition is a weakening process that causes consonants to become less occluded and more sonorous, or vowel-like, in some sense. It can be manifested in the voicing of voiceless consonants and in a more open articulation of consonants in general. Additionally, this weakening process can also cause for shorter duration of consonants, the extreme of which is deletion.

1.2. Dominican Spanish

The analysis of /s/ lenition in Dominican Spanish has the potential to elucidate whether high- or higher-frequency words can form part of the access route of retrieval of related low- or lower-frequency forms during production. In general, /s/ lenition in Spanish has been characterized in the literature as either aspiration or deletion (for example, *estás* > *e[h]tá[ø]* '2s to be'). Aspiration is caused by an opening of the articulation of the voiceless coronal sibilant [s] to a breathy pharyngeal approximant [h]. The result of the opening of this sibilant is more airflow, hence the breathiness of the sound, as well as a displacement of the place of articulation towards the back of the mouth. Deletion, as its name implies, results when no sound is present that might represent an underlying /s/.

Dominican Spanish displays very high rates of syllable- and word-final /s/ lenition. In fact, Lipski (1994:239) reports that /s/ lenition is nearly categorical even among educated speakers. Further, Lipski proposes that Dominican Spanish is unique among other high-reducing dialects in that syllable- and word-final /s/ in Dominican Spanish is deleted more often than aspirated. That is, rather than reducing the friction of the sibilant to a weak, breathy pharyngeal aspirated sound [h] (as in *e[h]tá[h]*), Dominican speech is known for lacking any sound that could represent an underlying /s/ (as in *e[ø]tá[ø]*).

Within a usage-based framework of phonology, the result of this rampant syllable- and word-final /s/ deletion is that the best exemplars of words which etymologically have this sound lack what would be considered an underlying /s/. (Terrell 1986 argues this point, but from a generative framework, that is, that an underlying /s/ in syllable-final position is now missing from the grammar of semi- and illiterate Dominicans.) To take a specific example, the majority of exemplars of *país* 'country' would not likely have four segments, as in /país/, but rather three /pai/. While there might not be categorical /s/ deletion in all exemplars of words with syllable- and word-final /s/, the vast majority of exemplars of words with word-final /s/ will likely have a deleted word-final /s/. Thus, all things being equal, the deleted /s/ version of *país* (that is, *paí*) will likely be selected during production.

1.3. Paradigmatic Peer-Pressure

Returning to our discussion of the influence of frequency differences between related words, it is proposed that high- or higher-frequency words are likely to form part of the access route of retrieval during the production of low- or lower-frequency words within the same paradigm as well as among derivational pairs. What makes this theory important for this paper is determining whether the deleted /s/ exemplar of words with word-final /s/ are part of the access route of retrieval of less-frequent inflectionally or derivationally related words in Dominican Spanish. If this is the case, then we would assume that the deleted word-final /s/ would likely have some effect on the word-medial /s/ in related forms. For example, the singular *país* 'country', which has a word-final /s/, is twice as

frequent as its plural counterpart *países*, which has a word-medial /s/. Consequently, we would expect the nearly-categorically-deleted word-final /s/ on the singular form to somehow affect the word-medial /s/ in the plural form. While we would not necessarily expect deletion of word-medial /s/ in syllable-initial position, as Spanish /s/ in this syllable position is not prone to deletion, or even aspiration, (see Brown 2005; File-Muriel and Brown 2010; Brown and Brown Forthcoming for exceptions), we would expect some kind of lenition of word-medial, syllable-initial /s/. This lenition may be manifested as temporal reduction, that is, a shortened duration, or with a reduced articulatory gesture, measurable as a lower frequency of friction (centroid), or as more voicing of the sibilant. These are the expected results when a word with word-final /s/ is more frequent than its word-medial-/s/ counterpart, as the higher-frequency word-final-/s/ word likely forms part of the access route of retrieval of the related lower-frequency word-medial-/s/ word. However, if the word-final /s/ word is of lower frequency than its counterpart word with word-medial /s/, then we would not expect such lenition of the word-medial /s/, as the higher-frequency word-medial /s/ word would likely be accessed directly (whole-form route) rather than by way of the related lower-frequency word with a word-final /s/ (compositional route).

Figure 2 represents in graphic form the proposed interaction of related words. The left side of the graphic shows the pair *páis-países*, in which the word-final-/s/ word *páis* is more frequent, represented by a larger font size and a larger and thicker circle. The influence of this higher-frequency word-final-/s/ word on the lower-frequency word-medial-/s/ word is also represented by a thicker arrow pointing towards its counterpart.

On the right side of the figure, the opposite scenario is presented. The singular word with word-final /s/, *mes* 'month', is about half as frequent as its plural counterpart *meses*. As such, the influence of the singular word with word-final /s/ on the plural word with word-medial /s/ should be smaller than in the other pair (*páis-países*). Again, the relative strength of the influence of the word-final-/s/ word on its word-medial-/s/ counterpart is represented by font size and line thickness. The result of this difference in relative strength of influence of the singular form on its corresponding plural form is that the plural word that is more influenced by its singular counterpart should have a more-lenited word-medial /s/ than the plural word that is less affected by its corresponding singular. In short, we would expect the word-medial /s/ of *países* to be more lenited than the word-medial /s/ of *meses*.

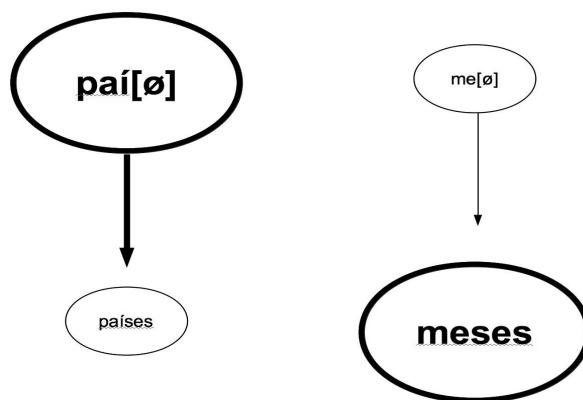


Figure 2: Influence of words with word-final /s/ on related words with word-medial /s/

Intuitively, this concept should apply not only to inflected pairs but to derived ones too. Thus, more-frequent base words should have more influence on their less-frequent derived forms, while base forms that are less frequent than their derived counterparts should have less of an effect on their counterparts, since those higher-frequency derived forms can be accessed whole, as per the dual-route model (Hay 2001; Baayen, Dijkstra, and Schreuder 1997; Bybee 1985; Eddington and Lestrade 2002). For example, the base word *diez* 'ten' is more frequent than its derived form *dieciocho* 'eighteen', while the base word *capaz* 'able' is less frequent than its derived form *capacidad* 'ability'. Thus, we would expect more influence from *diez* on *dieciocho* than we would expect from *capaz* on *capacidad*.

In addition to the influence of word-final-/s/ words on their word-medial-/s/ inflectionally or derivationally related counterparts, this paper also analyzes whether the reverse is true: Can more-frequent words with word-medial /s/ form part of the access route of related less-frequent words with word-final /s/? Taking the same examples presented above, can the more-frequent plural *meses* or the more-frequent derived noun *capacidad* form part of the access route of their related less-frequent words: singular *mes* and adjectival *capaz*? If so, we would expect the word-final /s/ on singular *mes* and on the adjective *capaz* to be more robust than would otherwise be the case, as the word-medial /s/ of *meses* and *capacidad* are in syllable-initial position and, therefore, are very likely to be maintained. Contrarily, the plural word *paises* is only half as frequent as its singular counterpart *pais*, and thus, we would expect *paises* to have little influence on the singular *pais*. More specifically, we would expect the word-final /s/ on the singular *pais* to be more lenited than the word-final /s/ on *mes* precisely because the plural counterpart of *pais* is comparatively less frequent than is the plural counterpart of *mes*. In graphic form, this idea would be expressed by reversing the direction and the thickness of the arrows in Figure 2.

1.4. Research Questions and Hypotheses

Assuming that the assertion of usage-based phonology is correct, that high- or higher-frequency words can form part of the access route of retrieval of related low- or lower-frequency words during production, we would expect the highly-lenited word-final /s/ in Dominican Spanish to cause more lenition of word-medial /s/ in a related word when the word-final-/s/ word is accessed during the production of the word-medial-/s/ word. In order to operationalize whether the word-final-/s/ word is accessed during the production of a related word-medial-/s/ word, a relative frequency ratio between the two words based on their respective lexical frequencies is created, as detailed in Section 2 below. In short, it is assumed that if a word with word-final /s/ is more frequent than its related word with word-medial /s/, then the word-final-/s/ word will likely be accessed during the production of the word-medial-/s/ word, and therefore, the compositional route of access will be taken. As such, the word-final-/s/ word will have a higher probability of influencing the phonetic realization of the word-medial-/s/ word and therefore will cause a more-lenited word-medial /s/. However, if a word with word-medial /s/ is more frequent than its related word with word-final /s/, then the assumption is made that the word with word-final /s/ will less likely be accessed, and therefore, the whole-form route of access of the word-medial-/s/ word will be taken. Consequently, the word-final-/s/ word will have a lower likelihood of affecting the realization of the word with word-medial /s/, and therefore, the word-medial /s/ should be more robust (less-lenited) than would be the case if its production were affected by the highly lenited version. Hence, this paper is based on the assumption that the phonetic realization of /s/ in Dominican Spanish can give insight into which access route of retrieval is employed during production.

This paper seeks to answer the following research questions:

- (1) Does the analysis of /s/ lenition in Dominican Spanish provide phonetic support for the idea of usage-based phonology that higher-frequency words can form part of the access route of retrieval of inflectionally and derivationally related lower-frequency words?
- (2) More specifically, does the analysis of the duration, the percentage of voicelessness, and the centroid of word-medial /s/ suggest that higher-frequency words with word-final /s/ can form part of the access route of retrieval during the production of related lower-frequency words with word-medial /s/?
- (3) Conversely, does the analysis of the duration, the percentage of voicelessness, and the centroid of word-final /s/ suggest that higher-frequency words with word-medial /s/ can form part of the access route of retrieval during the production of related lower-frequency words with word-final /s/?

It is hypothesized that the results of the statistical models will show that Dominican Spanish provides support that, indeed, higher-frequency words can form part of the access route of retrieval during the production of related lower-frequency words, whether the higher-frequency word contains a word-final /s/ and the related lower-frequency word contains a word-medial /s/, or vice versa.

As will be discussed in Section 3, the results provide support for the idea that higher-frequency words can be part of the access route of related lower-frequency words, but only in the analysis of word-medial /s/; no evidence for this assertion was found in the analysis of word-final /s/.

2. Methods

In order to test the assertion of usage-based phonology that higher-frequency words can form part of the access route of retrieval of related lower-frequency words, the duration in milliseconds, the percentage of voicelessness, and the centroid of /s/ in hertz was measured in a corpus of 85 spoken personal narratives of Dominican speakers native to and living in the northern city of Santiago (Chiara 1999). The corpus was recorded by two graduate students of Brigham Young University in 1999. The researchers created the corpus based on sociolinguistic factors such as education and socioeconomic class in order to obtain a well-balanced sample of spontaneous Dominican speech. The recordings were made with a digital audio tape recorder and a high-fidelity microphone and were conducted almost entirely in the homes of the informants. The researchers were mindful of trying to eliminate as much as possible the background noise that any home has. Nevertheless, there is occasional background noise in many recordings. The narratives were transcribed by various Spanish-speaking undergraduate students of the same university during subsequent years.

A list of related pairs that have a word-final /s/ with a corresponding word-medial /s/ was searched for within the transcriptions of 71 speakers of the corpus using the programming environment R (R Development Core Team 2010). This search returned 475 tokens from 29 different word types. The corresponding sound files were accessed with the phonetics software Praat (Boersma and Weenink 2010). Tokens that were followed by /s/ were excluded (N = 10), as it is extremely difficult, if not impossible, to distinguish where the first /s/ ends and the second one begins. Also, tokens of /s/ that were extraordinarily long, those in excess of 300 milliseconds, were excluded (N = 3). This left 462 tokens of both word-final and -medial /s/ from 28 word types for the analysis.

Both the waveform and the spectrogram of the interviews were viewed in order to manually delimit each token of /s/ in a TextGrid within Praat. A script then automated the measurement of the three dependent variables: (1) duration in milliseconds, (2) percentage of voicelessness, and (3) centroid in hertz. The duration was taken from the manually placed boundaries of /s/. In order to accurately measure the percentage of voicelessness and the centroid, tokens with background noise were excluded (N = 159), as noise skews these two measurements. The percentage of voicelessness was taken from Praat's Voice Report, which is based on the number of glottal pulses in a given interval, which in this case was the manually delimited boundaries of each token of /s/. To retrieve the centroid, the lowest 750 hertz were excluded from the signal with a Pass Hann filter, as voicing can skew the accuracy of the centroid measurement. Additionally, only the middle 60% of the manually delimited duration of /s/ was analyzed for centroid so as to avoid interference from the surrounding segments.

The predictor variables analyzed were: relative frequency, absolute frequency, rate of speech, preceding and following phonological contexts, prosodic stress, the morphological relationship between the token word and its related word, and the level of education of the speaker. The relative frequency was obtained by dividing the absolute frequency of the token word by the absolute frequency of the related word. Deciding which possible word would be considered the related one was based on absolute frequency, such that the most frequent of all the related words was taken to be the related one. For example, the related word of *capaz* 'able' was taken to be *capacidad* 'ability' rather than *capaces* 'able (pl)' because the former is more frequent than the latter. This decision was made despite the fact that sometimes a derived word was chosen as the related word rather than an inflected word, as with *capaz* and *capacidad*. The assumption is made that the most frequent of the related words has the highest resting activation state in memory (Bybee 2001, Hay 2001), and thus is the most likely word to have an effect on a neighboring related word during production. Basing "relatedness" on an empirical metric, such as absolute frequency, avoids any arbitrariness that might be present were the decision process based only on phonological and semantic similarities, notwithstanding the great importance of these similarities.

In order to control for variation in lexical frequencies across Spanish-speaking countries, the absolute frequency was taken from the corpus from which the tokens of this study were extracted. This corpus contains 278,949 words. However, in order to ensure that the frequency numbers obtained from this corpus are comparable to those found throughout the Spanish-speaking world, the frequency numbers were compared to those found in a much larger pan-Hispanic corpus of transcribed spontaneous speech in: Argentina (Barrenechea 1987), Bolivia (Mendoza 1996), Chile (Rabanales and Contreras 1979), Colombia (Instituto Caro y Cuervo 1986; Travis 2005; File-Muriel 2009; File-Muriel and Brown 2011), Costa Rica (Salguero 1976), Cuba (González Mafud 1996), Dominican Republic (Chiara 1999), Mexico (Lope Blanch 1971, 1976), Peru (Caravedo 1989), Puerto Rico (Amparo Morales 1990; Cortés-Torres 2005), Spain (Esgueva and Cantarero 1981; Marcos Marín 1992), Venezuela (Rosenblat 1979; Domínguez 1996), and the Southwestern United States (Lope Blanch 1990; Bills and Vigil 1999). The number of words in this larger corpus is 6,431,112. The frequency numbers obtained from these two corpora were compared with a Kendall's rank correlation tau test (rather than a Pearson's r test, as the frequency data are not distributed normally, as measured in two Shapiro-Wilk normality tests: Dominican corpus, $W = 0.77$, $p \leq 0.001$; pan-Hispanic corpus, $W = 0.89$, $p \leq 0.001$) and found to be very highly correlated: $\tau = 0.95$, $z = 27.13$, $p \leq 0.001$. Given this strong correlation, only the frequency numbers from the Dominican corpus were included in the statistical models reported below.

The rate of speech was measured by taking the number of segments in the three-word string immediately around each token and dividing this number by the duration of this three-word string. This created an index of segments per second immediately around each token. The /s/ segment and its duration were removed from this calculation so as to avoid circularity. When the word with the current token of /s/ was either preceded or followed by a pause, the measurement was accordingly reduced so that the rate of speech of some tokens was calculated with fewer than three words. The advantage of using a localized rate of speech measurement for each token is to control for natural fluctuations in rate of speech during spontaneous speech.

Brown (2005) shows that preceding non-high vowels favor lenition of /s/ in New Mexican Spanish. Because of this finding, the preceding phonological context was coded as either a high vowel /i, u/ or a non-high vowel /a, e, o/. The following phonological context was coded similarly, but with more categories: high-vowel, non-high vowel, coronal consonant, non-coronal consonant, and pause.

The prosodic stress was coded simply as atonic and tonic, assuming resyllabification when word-final tokens were followed by a vowel.

The morphological relationship between the token word and its related word was coded as either inflectional or derivational. Bybee (1985) argues that words are more closely related to other words when they share more semantic similarities than similarities in form. As such, words of the same inflectional paradigm are likely more closely related than words related by derivation, as derivation implies a change, albeit slight, in lexical meaning.

The educational level of the speakers was measured by placing speakers on an ordinal index in which a given speaker had more years of educational training than a speaker one index step below and fewer years of education than a speaker one step above. However, these index steps did not necessarily correspond to a one-to-one increase in years of educational training, so that a given speaker could have had two or three more years of education than a speaker one index step below, but only one year less than a speaker one step above. The educational training levels range from some elementary school to medical school graduate. It is hypothesized that speakers with more education and thus more familiarity with orthography, especially the orthography of less-frequent words, will be more affected by this orthography and thus more prone to more fully pronounce /s/.

In order to statistically control for the effect of all predictor variables and thus analyze the influence of the relative frequency of a given token word in comparison to its related form, a series of stepwise linear regressions were fitted with the statistics software R (R Development Core Team 2010). Six regressions in total were run: a separate model was fitted for each of the three dependent variables within the two word positions (medial and final).

3. Results

The results of the linear regressions of the duration and the percentage of voicelessness of word-medial /s/ lend support to the usage-based idea that higher-frequency words can be part of the access route of retrieval during the production of related lower-frequency forms. However, no evidence of support for the theory is seen in the analysis of the centroid of word-medial /s/ nor in any of the three dependent variables in words with /s/ in word-final position.

3.1. Word-medial /s/

Significant influence from the relative frequency is found in the linear regressions of the duration and the voicelessness of /s/, but not in the centroid.

3.1.1. Duration

The results of the linear regression show that the duration of word-medial /s/ is conditioned by: rate of speech, relative frequency, prosodic stress, and absolute frequency. See Table 1.

Table 1: Linear regression of duration of word-medial /s/

Predictor	Coefficient	Std error	T-value	P-value
(Intercept)	0.126	0.012	10.119	0.000
Rate of speech	-0.004	0.001	-6.957	0.000
Relative frequency	0.009	0.003	3.290	0.001
Stress = tonic	0.030	0.011	2.588	0.010
Log absolute freq	0.004	0.002	2.251	0.026

N = 179; adjusted R² = 0.233; F = 11.83; df = 5, 173; p-value ≤ 0.001

As expected, word-medial /s/ duration is inversely related to rate of speech, indicated by the negative coefficient. In other words, duration decreases as rate of speech increases.

Second only to rate of speech in the hierarchy of predictors is the relative frequency. With a positive coefficient returned by the linear regression, a positive correlation exists between relative frequency and duration of word-medial /s/. As words with word-medial /s/ become more frequent in comparison to their related word with word-final /s/, the duration of word-medial /s/ increases, as predicted within usage-based phonology.

Third, stress was also selected as having a significant effect on word-medial /s/ duration. In comparison to atonic syllables (which are absent from the table, as they are the baseline level against which the tonic syllables are measured), tonic syllables prefer longer durations of word-medial /s/.

Fourth and finally, the duration of word-medial /s/ is also conditioned by the absolute frequency of the word in which /s/ occurs. Duration of /s/ lengthens as the absolute frequency increases.

3.1.2. Percentage of voicelessness

The results of the linear regression show that the percentage of voicelessness of word-medial /s/ is conditioned by the absolute frequency and the relative frequency of the word. See Table 2.

Table 2: Linear regression of percentage of voicelessness of word-medial /s/

Predictor	Coefficient	Std error	T-value	P-value
(Intercept)	0.901	0.063	14.336	0.000
Log absolute freq	-0.050	0.013	-3.723	0.000
Relative frequency	0.100	0.038	2.608	0.010

N = 106; adjusted R² = 0.113; F = 7.718; df = 2, 103; p ≤ 0.001

The percentage of voicelessness of /s/ is inversely correlated to the absolute frequency of the word in which /s/ occurs, such that as the word frequency increases, the amount of voicelessness decreases, that is, the amount of voicing increases.

The second, and only other significant, predictor of voicelessness is the relative frequency of the word in which word-medial /s/ occurs in comparison to its related word. As expected, as words become more frequent than their related words, the amount of voicelessness of word-medial /s/ increases, or in other words, the amount of voicing decreases.

3.1.3. Centroid

The centroid was not affected by the relative frequency of the word in which /s/ occurs. In the end, the only predictor that significantly conditions the centroid is the morphological relation that words with word-medial /s/ have with their related words. In comparison to derivationally related words, inflectionally related words have a word-medial /s/ with a higher centroid.

3.2. Word-final /s/

As mentioned above, there is no evidence from the analysis of word-final /s/ that suggests a higher-frequency word with word-medial /s/ can form part of the access route of a related lower-frequency word with word-final /s/. The relative frequency of a word with word-final /s/ in comparison to its related word with word-medial /s/ was not selected as making a significant contribution to the prediction of any of the three dependent variables in word-final position.

3.2.1. Duration

The results of the linear regression show that the duration of word-final /s/ is conditioned by: rate of speech, educational level of the speaker, the following phonological context, and the absolute frequency of the word with word-final /s/. See Table 3.

Table 3: Linear regression of duration of word-final /s/

Predictor	Coefficient	Std error	T-value	P-value
(Intercept)	0.155	0.023	6.627	0.000
Rate of speech	-0.005	0.001	-5.745	0.000
Educational level	0.007	0.001	5.987	0.000
Following segment = non-high V	-0.041	0.012	-3.553	0.000
Following segment = non-coronal C	-0.041	0.009	-4.508	0.000
Following segment = high V	-0.061	0.026	-2.345	0.020
Following segment = coronal C	-0.045	0.012	-3.832	0.000
Log absolute frequency	-0.010	0.003	-3.067	0.002

N = 283; adjusted R² = 0.425; F = 24.19; df = 9, 273; p ≤ 0.001

As expected, and similar to the results in word-medial position, word-final /s/ is shortened as the rate of speech increases.

The second most influential predictor is the educational level of the speaker. Speakers with more educational training produce longer word-final /s/ than those with less education. This result concurs with previous studies (for example, Alba 2004) that detail the effect of education on syllable- and word-final /s/ lenition in Spanish.

The third predictor variable to condition word-final /s/ is the following phonological context. In comparison to a following pause, all other sound groups (high vowel, non-high vowel, coronal consonant, non-coronal consonant) significantly condition a shorter duration of word-final /s/.

Lastly, the absolute frequency of the words with word-final /s/ conditions the duration of this sound. As expected, as absolute frequency increases, word-final /s/ duration decreases.

3.2.2. Percentage of voicelessness

Similar results were found in the linear regression of the percentage of voicelessness of word-final /s/. The predictors selected as making a significant contribution to the prediction of the percentage of voicelessness are: the rate of speech, the preceding phonological context, the educational level of the speaker, the morphological relation of the token word and its related word, and the following phonological context.

Table 4: Linear regression of percentage of voicelessness of word-final /s/

Predictor	Coefficient	Std error	T-value	P-value
(Intercept)	0.883	0.151	5.860	0.000
Rate of speech	-0.022	0.007	-3.198	0.002
Preceding context = high V	0.206	0.061	3.385	0.001
Educational level	0.029	0.009	3.212	0.002
Morphological relation = inflection	-0.335	0.092	-3.649	0.000
Following context = non-high V	-0.082	0.088	-0.937	0.350
Following context = non-coronal C	-0.291	0.073	-3.970	0.000
Following context = high V	-0.266	0.187	-1.427	0.155
Following context = coronal C	-0.217	0.097	-2.246	0.026

N = 197; adjusted R² = 0.300; F = 11.49; df = 8, 188; p ≤ 0.001

The predictor that most conditions the percentage of voicelessness is the rate of speech. As expected, an increase in rate of speech causes a decrease in voicelessness, or in other words, an increase in voicing, indicative of lenition.

The second most influential predictor is the preceding phonological context. In comparison to preceding high vowels, non-high vowels increase the percentage of voicing of word-final /s/.

Third, the educational level of the speaker makes a significant contribution to the prediction of the percentage of voicelessness of word-final /s/. As educational level increases, so does the amount of voicelessness.

Fourth, the morphological relation between the words with word-final /s/ and their related words significantly conditions the percentage of voicelessness. Derivation influences voicing more than inflection does in these data.

Fifth and finally, the following phonological context significantly conditions word-final /s/ voicing. In comparison to a following pause, a following consonant significantly influences voicing of word-final /s/. However, the effect of a following vowel does not significantly differ from the effect of a following pause.

3.2.3. Centroid

As with the previous linear regressions on word-final /s/, the regression model of the centroid of word-final /s/ failed to return relative frequency as a significant predictor. The significant predictors are: rate of speech, educational level, and absolute frequency.

Table 5: Linear regression of centroid of word-final /s/

Predictor	Coefficient	Std error	T-value	P-value
(Intercept)	7018.722	1422.314	4.935	0.000
Rate of speech	-235.255	49.330	-4.769	0.000
Educational level	221.245	72.148	3.067	0.002
Log absolute frequency	-418.402	196.720	-2.127	0.035
Preceding context = high V	661.489	470.542	1.406	0.161

N = 197; adjusted R² = 0.212; F = 14.16; df = 4, 192; p ≤ 0.001

First, the rate of speech is the most influential predictor of word-final /s/ centroid. As expected, an increase in rate of speech brings with it a decrease in centroid, indicative of lenition.

The educational level of the speaker also significantly conditions word-final /s/ centroid. Speakers with more education have a higher centroid, indicating less lenition of word-final /s/.

Last, the absolute frequency was also selected as a significant predictor. Absolute frequency has an inverse relation with centroid, such that an increase in absolute frequency causes a decrease in centroid.

4. Discussion

The results of the linear regressions are mixed in their assessment of the influence of related words on the realization of /s/ in Dominican Spanish. While there is no effect on word-final /s/, two of the three linear regressions of word-medial /s/ found a significant conditioning effect from the relative frequency of a given word in comparison to its related word. Both the duration and the percentage of voicelessness of word-medial /s/ were significantly conditioned by relative frequency. These results support the argument of usage-based phonology (Bybee 2001:116) that high-frequency words can be part of the access route of retrieval of low-frequency forms in the same paradigm. Further, these results lend credence to the version of the dual-route model (proposed by Bybee (1985) and Baayen, Dijkstra, and Schreuder (1997), among others) that proposes that higher-frequency derived words need not be parsed by means of their lower-frequency base forms.

These data answer affirmatively the first research question posed in this paper: the analysis of /s/ lenition in Dominican Spanish provides evidence, albeit limited, that higher-frequency words can be part of the access route of inflectionally and derivationally related lower-frequency words. Further, the second research question is answered affirmatively: higher-frequency words with word-final /s/ can be part of the access route of related lower-frequency words with word-medial /s/, at least when measured in the duration and percentage of voicelessness of word-medial /s/. However, the third research question is answered negatively by these data: there is no evidence to suggest that higher-frequency words with word-medial /s/ form part of the access route of related lower-frequency words with word-final /s/. This lack of a significant conditioning effect from relative frequency in word-final position needs to be studied in future research. One possible explanation may relate to the fact that the singular is most commonly the unmarked member of a singular-plural paradigm and the majority of the word-final /s/ tokens were singular /s/. Thus, the unmarked member, which is usually the most frequent member of a paradigm anyway (see Tiersma (1982) for a discussion of the influence of frequency on markedness), may be more insulated from the influence of other members of the paradigm. Further study will have to address such a possibility. Additionally, the fact that word-final /s/ is overwhelming in syllable-final position, and is therefore overwhelming deleted in Dominican Spanish, may create a ceiling effect on the influence of relative frequency on this sound.

In addition to providing support for the usage-based proposal that higher-frequency words can be part of the access route of retrieval of related lower-frequency words, these data provide support, and in some cases strong support, for the significant effect of several predictor variables shown in the literature to condition /s/ lenition in Spanish. The results of the linear regressions show that the rate of speech is a strong predictor of /s/ lenition, having been selected in four of the six models and having been selected as the most influential variable in all of those models. It is intuitive that rate of speech would influence /s/ duration because, simply put, the duration of each segment should decrease as the rate of speech increases. The definition of an increase in rate of speech is just that: to pronounce the same number of segments in less time. However, the fact that rate of speech was also selected in the models that analyze percentage of voicelessness and centroid is important. Not only does an increase in the rate of speech have a leniting effect on /s/ duration, it also contributes to /s/ lenition by increasing /s/ voicing and decreasing the centroid of /s/. These results fall in line with those of a recent study on both syllable-initial and syllable-final /s/ lenition in Cali, Colombia (File-Muriel and Brown 2010, 2011).

Likewise, the results of this paper concur with previous accounts (among them, Terrell (1979),

Brown (2004), and File-Muriel and Brown (2010, 2011)) in demonstrating that the preceding and following phonological contexts as well as the absolute frequency of a word with /s/ are also significant predictors of /s/ lenition in Spanish.

In summary, this paper provides support, albeit limited, for the argument of usage-based phonology that higher-frequency words can form part of the access route of retrieval during the production of related lower-frequency words with an analysis of /s/ lenition in Dominican Spanish. This analysis of spontaneous speech in personal narratives contributes to the results of the (mostly) experimental data reported in previous studies, especially those that address the dual-route model.

References

- Alba, Orlando. 2004. *Cómo Hablamos Los Dominicanos: Un Enfoque Sociolingüístico*. Santo Domingo, Dominican Republic: Grupo León Jimenes.
- Amparo Morales, María Vaquero. 1990. *El habla culta de San Juan*. Río Piedras: Editorial de la Universidad de Puerto Rico.
- Baayen, R. Harald, Ton Dijkstra, and Robert Schreuder. 1997. Singulars and Plurals in Dutch: Evidence for a Parallel Dual-Route Model. *Journal of Memory and Language* 37.94-117.
- Barrenechea, Ana María. 1987. *El habla culta de la ciudad de Buenos Aires: materiales para su estudio*. Buenos Aires: Universidad Nacional de Buenos Aires.
- Bills, Garland D., and Neddy A. Vigil. 1999. Ashes to ashes: The historical basis for dialect variation in New Mexican Spanish. *Romance Philology* 53.43-67.
- Boersma, Paul, and David Weenink. 2010. *Praat: doing phonetics by computer*. www.praat.org.
- Browman, Catherine P., and Louis Goldstein. 1992. Articulatory Phonology: An Overview. *Phonetica: International Journal of Speech Science* 49.155.
- Brown, Earl K., and Esther L. Brown. Forthcoming. Syllable-final and syllable-initial /s/ reduction in Cali, Colombia: One variable or two? *Colombian Varieties of Spanish*, ed. by Richard File-Muriel and Rafael Orozco. Madrid: Iberoamericana.
- Brown, Esther L. 2004. The reduction of syllable initial /s/ in the Spanish of New Mexico and Southern Colorado: A usage-based approach. PhD thesis. Albuquerque: University of New Mexico.
- 2005. New Mexican Spanish: Insight into the Variable Reduction of “la ehe inihial” (/s-/). *Hispania: A Journal Devoted to the Teaching of Spanish and Portuguese* 88.813-824.
- Bybee, Joan. 1985. *Morphology: A Study of the Relation Between Meaning and Form*. Typological studies in language, v. 9. Philadelphia: John Benjamins.
- 2001. *Phonology and language use*. Cambridge Studies in Linguistics 94. Cambridge: Cambridge University Press.
- 2006. From Usage to Grammar: The Mind’s Response to Repetition. *Language* 82.711-733.
- 2010. *Language, Usage and Cognition*. Cambridge University Press.
- Caravedo, Rocio. 1989. *El español de Lima: materiales para el estudio del habla culta*. Lima: Pontificia Universidad Católica del Perú.
- Chiara, Angela O. 1999. Analysis and background information for informants: Santiago, Dominican Republic linguistic corpus. MA thesis. Provo, Utah: Brigham Young University.
- Cortés-Torres, Mayra. 2005. La perífrasis *estar* + *-ndo* en el español puertorriqueño: variación dialectal o contacto lingüístico? University of New Mexico.
- Davies, Mark. 2009. The 385+ Million Word Corpus of Contemporary American English (1990-2008+): Design, Architecture, and Linguistic Insights. *International Journal of Corpus Linguistics* 14.159-190.
- Domínguez, Carmen Luisa. 1996. El habla de Mérida: un corpus de estudio. *Lengua y Habla* 1.46-55.
- Dykstra, G. 1955. Spectrographic analysis of Spanish sibilants and its relation to Navarro’s physiological phonetic descriptions. PhD thesis. Ann Arbor: University of Michigan.
- Eddington, David, and Patricia Lestrade. 2002. Are Plurals Derived or Stored? *Structure, Meaning, and Acquisition in Spanish: Papers from the 4th Hispanic Linguistics Symposium*, 269-284. Somerville, MA: Cascadilla Press. <http://www.cascadilla.com/hls4.html>.
- Esgueva, Manuel, and Margarita Cantarero. 1981. *El habla de la ciudad de Madrid: materiales para su estudio*. Madrid: Instituto “Miguel de Cervantes.”
- File-Muriel, Richard, and Earl K. Brown. 2010. The gradient nature of /s/-lenition in Caleño Spanish. *University of Pennsylvania Working Papers in Linguistics* 16. <http://repository.upenn.edu/pwpl/vol16/iss2/7>.
- 2011. The gradient nature of s-lenition in Caleño Spanish. *Language Variation and Change* 23.
- File-Muriel, Richard. 2009. The role of lexical frequency in the weakening of syllable-final lexical /s/ in the Spanish of Barranquilla, Colombia. *Hispania* 92.348-360.

- González Mafud, Ana María. 1996. La Habana: Segunda época. *Memoria de la V reunión de trabajo de la comisión ejecutiva del estudio del español hablado culto*, ed. by J. M. Lope Blanch, 87-94. Universidad Nacional Autónoma de México.
- Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39.1041-1070.
- Instituto Caro y Cuervo. 1986. *El habla de la ciudad de Bogotá: materiales para su estudio*. Bogotá: Publicaciones del Instituto Caro y Cuervo.
- Lewis, Anthony M. 2001. Weakening of intervocalic /p, t, k/ in two Spanish dialects: Toward the quantification of lenition processes. University of Illinois.
- Lipski, John. 1994. *Latin American Spanish*. New York: Longman.
- Lope Blanch, Juan M. 1971. *El habla de la Ciudad de México: materiales para su estudio*. México, D. F.: Universidad Nacional Autónoma de México.
- 1976. *El habla popular de la ciudad de México: materiales para su estudio*. México, D. F.: Universidad Nacional Autónoma de México.
- 1990. *El español hablado en el suroeste de los Estados Unidos: materiales para su estudio*. México, D. F.: Universidad Nacional Autónoma de México.
- Marcos Marín, Francisco. 1992. *Corpus oral de referencia de la lengua española contemporánea*. Madrid: Universidad Autónoma de Madrid. http://elvira.llf.uam.es/does_es/corpus/corpus.html.
- Mendoza, Jose G. 1996. *Léxico del habla culta de La Paz*. La Paz, Bolivia: Universidad Mayor de San Andrés.
- Pharies, David A. 2007. *Breve historia de la lengua española*. Chicago: University of Chicago Press.
- Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Frequency and the Emergence of Linguistic Structure*, ed. by Joan Bybee and Paul Hopper, 137-157. Philadelphia: John Benjamins Publishing Company.
- 2003. Probabilistic phonology: Discrimination and robustness. *Probabilistic Linguistics*, ed. by R. Bod, Jennifer Hay, and S. Jannedy, 177-228. Cambridge, Mass.: The MIT Press.
- Pinker, Stephen, and Alan Prince. 1991. Regular and Irregular Morphology and the Psychological Status of Rules of Grammar. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society February 15-18, 1991: General Session and Parasession on the Grammar of Event Structure*, 230-251. Berkeley: Berkeley Linguistics Society.
- R Development Core Team. 2010. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org>.
- Rabanales, Ambrosio, and Lidia Contreras. 1979. *El habla culta de la ciudad de Santiago: materiales para su estudio*. Santiago de Chile: Editorial Universitaria.
- Rosenblat, Angel. 1979. *El habla culta de Caracas: materiales para su estudio*. Caracas: Universidad Central de Venezuela.
- Salguero, Miguel. 1976. *Así vivimos los Ticos*. San José: Editorial Universitaria Centroamericana.
- Salvador, Gregorio. 1968. Neutralización de G-/K- en español. *Actas del XI Congreso Internacional de Lingüística y Filología Románicas*, 4:1739-1752. Madrid.
- Terrell, Tracy D. 1979. Final /s/ in Cuban Spanish. *Hispania* 62.599-612.
- 1986. La desaparición de /s/ posnuclear al nivel léxico en el habla dominicana. *Estudios sobre la fonología del español del Caribe*, 117-134. Caracas: Ediciones La Casa de Bello.
- Tiersma, Peter Meijes. 1982. Local and General Markedness. *Language* 58.832-849.
- Travis, Catherine E. 2005. *Discourse Markers In Colombian Spanish: A Study In Polysemy*. Cognitive Linguistics Research. Berlin: Mouton de Gruyter.

Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology

edited by Scott M. Alvord

Cascadilla Proceedings Project Somerville, MA 2011

Copyright information

Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology
© 2011 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-449-2 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Brown, Earl K. 2011. Paradigmatic Peer-Pressure: Word-Medial, Syllable-Initial /s/ Lenition in Dominican Spanish. In *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology*, ed. Scott M. Alvord, 46-58. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2634.