Of Categories and Continua: Relating Discrete and Gradient Properties of Sociophonetic Variation

Daniel Erker*

1 Introduction

Sociophonetic variation is typically examined from either a categorical or a continuous perspective, but usually not both in the same analysis. Thus variable realizations of coda /t, d/ and /r/ in English are typically treated discretely in terms of the phonetic presence or absence of these phonological units, that is in terms of *t/d-deletion* (Guy 1980, Labov et al. 1968) or *r-lessness* (Labov 1966). Similarly, variability in the realization of Spanish coda /s/ is often described in terms of three discrete values, [s], [h], or [Ø] (Bybee 2010, Cedergren 1973, Lipski 1985). Vowels, however, have been extensively studied from a continuous perspective that examines differences in gradient acoustic parameters such as formant values (Labov 1994) or duration (Scobbie et al. 1999). While some variability in consonant production has been explored from a continuous perspective (Foulkes et al. 2006) and several studies of vowel production utilize discrete classifications (Watt and Milroy 1999), it remains rare to find discrete and continuous methods of analysis applied to the same phenomenon in the same study.

There is reason to be concerned about this trend, given that correlations between linguistic forms and social factors can be found at all levels of phonetic and phonological structure (Foulkes and Docherty 2006:412). The hypothesis investigated here is that models of sociophonetic variation that do not distinguish and relate discrete and continuous levels of expression may fail to identify significant patterns and restrict our understanding of linguistic and social conditioning factors. The current study explores this hypothesis through an examination of variation in the production of Spanish coda /s/.

Typically, variation in the production of Spanish coda /s/ is described in terms of three discrete variants: The first variant, represented by the segment [s], is used to describe cases in which phonological coda /s/ is produced as a voiceless alveolar fricative. That is to say, that the word *mismo,* 'same', is produced as [mismo]. Alternatively, /s/ may also be produced as a voiceless glottal fricative, represented segmentally as [h]. Here, /mismo/ is realized as [mihmo]. A third possibility is that in the phonetic implementation of a phonological representation that contains a coda /s/, there is, in fact, no evidence in the speech stream of any sound corresponding to that segment. Such cases are typically described as instances of *s-deletion*, and are regularly represented with the symbol [Ø]. The majority of research on /s/ variation employs this metric (Cedergren 1973, Lipski 1985, 1994).

The current study explores variability in /s/ production using a different method. Data are analyzed from a binary, discrete perspective, focusing on the alternation between the presence and absence of frication. In addition, tokens are also examined in terms of their spectrotemporal properties. Fricative moments are described in terms of two acoustic parameters: (1) duration in milliseconds and (2) mean *center of gravity* (COG) in Hertz, which is a measure that can be used to quantify the spectral distribution and amplitude of turbulent noise.

Data are examined within the context of language and dialect contact present in New York City (NYC). Results suggest that a unified methodology does more than simply increase the descriptive breadth of the analysis. Instead, it illustrates that certain patterns of variation are opaque at exclusively categorical or gradient levels. While some factors robustly condition /s/ production regardless of the method of analysis, others display predictive powers that are restricted to either the discrete or continuous domain of variability. Furthermore, the data reveal that categories and continua are equally important tools in assessing generational differences in /s/ production among Spanish speakers in NYC.

*I would like to thank the audiences at the Sociolinguistics Lunch Series at the Graduate Center of the City University of New York and at the NWAV 40 meeting for their insightful comments and questions.

2 The Study of Spanish in New York City

There are many questions for scholars of Spanish to choose from when examining the linguistic situation of New York City. Two of the most general are (1) How does Spanish spoken in NYC compare to Spanish spoken in Latin America? and (2) If there are differences between them, what are they, in what grammatical domains are they manifested, and what accounts for them? To answer these questions it is essential to establish a demographic profile of the Spanish speaking community in NYC. An imperfect but nonetheless valuable source of information for this task is Census data. Among the shortcomings of Census data is the fact that, due to the potential reluctance of undocumented residents to filling out Census forms, the size of the Hispanic population in NYC is likely to be underreported. Furthermore, the racial categories included in the Census are problematic in requiring that participants choose between potentially overlapping categories. For instance, the choice between *Black/African-American* and *Hispanic* may inadequately reflect the identities of New Yorkers of Afro-Caribbean origin or others who consider themselves Afro-Latinos. Despite these disadvantages, there are a number of reliable generalizations to be made on the basis of Census data that are useful in guiding linguistic analysis. Below, Table 1 reports the population, by race, of NYC (as constituted by the boroughs of Brooklyn, the Bronx, Manhattan, Queens, and Staten Island) in 1990, 2000, and 2010.

Total Population of New York City	1990		2000		2010	
	N	%	N	%	N	%
	7,322,564	100	8,008,278	100	8,175,133	100
White non Hispanic	3,163,125	43.2	2,801,267	35	2,722,904	33.3
Black/African American	1,847,049	25.2	1,962,154	24.5	1,861,295	22.8
Asian or Pacific Islander	489,851	6.7	783,058	9.8	1,030,914	12.6
American Indian	17,871	.2	17,321	.2	17,427	.2
Some Other Race	21,157	.3	58,775	.7	57,841	.7
Two or more races[1]	-	-	225,149	2.8	148,676	1.8
Hispanic Origin	***1,783,511***	***24.4***	***2,160,554***	***27***	***2,336,076***	***28.6***

Table 1: Total Population by Mutually Exclusive Race and Hispanic Origin. New York City and Boroughs, 1990 to 2010.

The table shows that as of 2010, nearly one third of New Yorkers were of Hispanic origin. It also shows that this group grew steadily and at a faster rate than the overall population of NYC between 1990 and 2010. In addition, Census data indicates that the Hispanic population of NYC, which has traditionally been characterized by a large Caribbean majority, has in recent decades experienced a steady increase in the number of Hispanics from Mainland Latin America. As of 2010, 82% of Hispanic New Yorkers had origins in one of six countries, three located in the Caribbean and three in the Mainland: Puerto Rico, Dominican Republic, Cuba, Mexico, Ecuador, and Colombia. Table 2 gives the figures associated with each of the countries.

Six largest Hispanic groups in NYC	2010	
	Country	N
Caribbean	Puerto Rico	723,621
	Dominican Republic	576,701
	Cuba	40,840
Latin American Mainland	Mexico	319,263
	Ecuador	167,209
	Colombia	94,723
Total		1,922357

Table 2: NYC Hispanic Population by selected subgroups. New York City and Boroughs, 2010.

[1]Data for this category was not reported until 2000.

To date, the largest scale effort to analyze Spanish spoken in the New York area is that represented by the *Otheguy Zentella Corpus of Spanish in New York City*, which was collected between 2000 and 2004. This corpus consists of 140 sociolinguistic interviews with speakers whose origins are in one of the six countries included in Table 2 above. Seventy-two speakers in the corpus have origins in the Caribbean (Puerto Rico, Dominican Republic, or Cuba), and sixty-eight speakers originate from the Latin America mainland (Mexico, Colombia, Ecuador).[2] In addition to variability in regional origin, speakers vary across a range of other parameters, including age, gender, and socioeconomic status, as well as age of arrival to and time spent living in NYC. With respect to the research questions presented at the beginning of this section, recent analyses suggest that Spanish spoken in New York City has much in common with that of Latin America, but that there is evidence of a growing discontinuity between them (Otheguy et al. 2007, Otheguy and Zentella 2012).

The primary findings of this research are related to variation in the use of subject personal pronouns. Variation in pronoun use has been widely studied in Spanish sociolinguistics, and numerous studies have demonstrated that the presence and absence of pronouns is significantly conditioned by linguistic and social factors (Bayley and Pease-Alvarez 1996, Cameron 1993, Guitart 1982, Silva-Corvalán 1982). The examples in (1) below illustrate the relevant variation.

(1) a. *yo canto* (pronoun present)
b. *canto* (pronoun absent)
'I sing'

A much replicated finding in the literature is that rates of pronoun presence are typically higher in Caribbean varieties of Spanish compared to those observed in Mainland regions (Otheguy and Zentella 2012).

Otheguy and Zentella report that Spanish speakers raised in NYC use pronouns in a different way than those who recently arrived from Latin America. *Recent arrivals* are those speakers who came to NYC after their eighteenth birthday and had spent less than six years in NYC before their interview. Speakers were considered to be *Raised in NYC* if they were born in New York or had been brought to NYC by their third birthday. Data reveal that overall rates of pronoun use are significantly higher in the speech of the NYC raised. Additionally, they show that regional differences in the pronoun use of recent arrivals are diminished among the NYC raised. That is, the tendency for Caribbean and Mainland speakers to differ significantly in their rates of pronoun use is attenuated in the second generation. Otheguy and Zentella interpret these generational differences as the result of two kinds of contact. Contact with English, itself a so-called *non-pro-drop* language, promotes an increase in rates of pronoun use among the NYC raised. In addition, the interaction of Caribbean and Mainland speakers in NYC promotes a process of mutual accommodation: "Both dialect and language contact are shaping Spanish in New York City and promoting, in the second generation, the formation of a New York Spanish speech community" (Otheguy, Zentella, and Livert 2007:770).

The current study draws its data from the corpus described above, investigating whether the generational discontinuity in morphosyntax reported by Otheguy and Zentella is paralleled in the phonological domain. Results reveal a generational shift in the variable production of coda /s/ that is similar to the trend observed in pronoun use, thus offering additional support to the notion of an emerging New York Spanish speech community. However, in order to properly assess the data for /s/ production, both discrete and continuous levels of analysis are required.

3 Data and Method

The present analysis considers the speech of twenty speakers included in the Otheguy Zentella Corpus. This group consists of ten recent arrivals and ten NYC raised speakers, with equal numbers of Mainlanders and Caribbeans across the generational groups. Two hundred cases of coda /s/

[2]While the *Otheguy Zentella* corpus includes a number of speakers from the coastal regions of Mexico, Colombia, and Ecuador, the Mainland speakers included in the current study all originate from the interiors of these countries.

were identified in the transcripts[3] of each speaker, for a total of 4,000 tokens. Data was collected at two different points. The first point was at ten minutes into a speaker's interview and the second at twenty minutes. The first 100 tokens of coda /s/ that occurred after each of those points were identified. All tokens were then subjected to acoustic analysis in Praat.

The presence of fricative moments was established on the basis of characteristic cues in spectrographic and waveform representations of speech. These include random noise located in the upper limits of the spectrogram and aperiodic waveforms in the mid-high frequency range. All cases illustrating acoustic moments typical of frication were coded as *frication present*. Those cases in which evidence of frication was absent from the speech stream were coded as *frication absent*. In 2,912 tokens, or 72.8 percent of the data, spectrographic and waveform evidence indicated the presence of speech-generated frication. In the other 1,088 cases, or 27.2 percent of the data, it did not. Figure 1 illustrates these data.

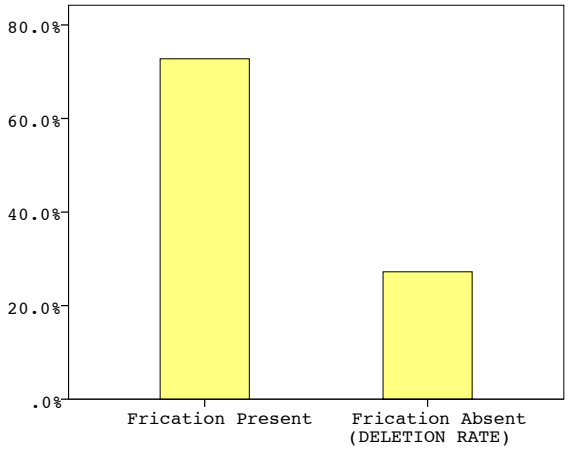

Figure 1: Percent presence and absence of frication.

In addition, all fricative moments were measured in terms of their duration in milliseconds and their mean center of gravity in Hertz. This process is illustrated in Figure 2 below, which shows a spectrogram for the utterance *tres hermanos,* 'three siblings'.

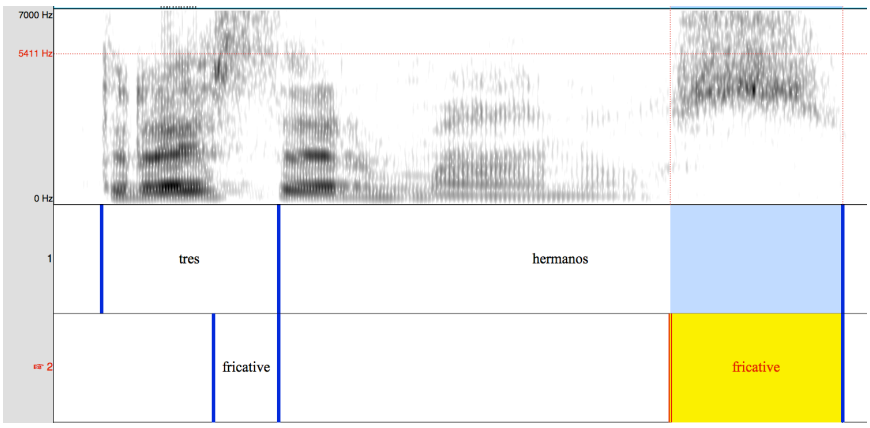

Figure 2: Spectrogram for two tokens of coda /s/, in *tres hermanos.*

In the first row below the spectrogram, *tres* and *hermanos* have each been bounded off in a Praat

[3]Interviews were transcribed by the research team that initially collected the data, not by the author of the present study.

textgrid. Below this row, two intervals of frication have been segmented and labeled *fricative.*[4] The interval of frication on the left is clearly shorter than that on the right (in yellow). Additionally, an examination of the spectrogram within each interval shows that the fricative noise of *tres* is distributed throughout the frequency spectrum, that is, there is sonic energy present at both the lower and higher frequencies. This contrasts with the turbulent noise at the end of *hermanos*, which is concentrated exclusively in the higher frequencies.

The temporal difference between these two fricative intervals can be captured by reporting their duration in milliseconds (ms). The spectral differences between these tokens, in particular the distribution and amplitude of turbulent noise, can be characterized with the parameter known as *center of gravity* (COG). COG is a weighted average calculated with the equation $COG = \sum fI / \sum I$ where I is the amplitude in decibels and f the frequency in Hertz of the spectral components. COG was measured at the one-quarter, one-half, and three-quarters points of each fricative interval. These three values were averaged, giving a mean COG for the token. Using these two measures, the two fricative intervals above have, respectively, durations of 104 and 206 ms and mean COGs of 2098 and 4821 Hz. Figure 3 plots duration and COG for study's dataset. Mean frication duration in the sample is 55.9 ms, and mean COG is 1626 Hz.

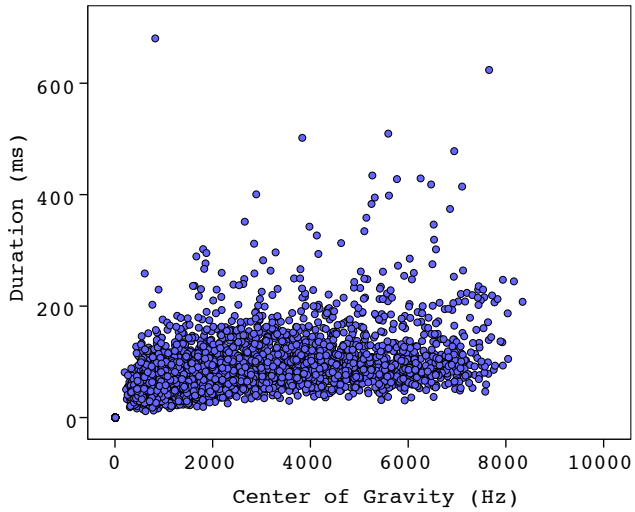

Figure 3: Duration and mean COG for 4,000 tokens of coda /s/.

4 Results

Returning briefly to subject pronouns, analysis of the twenty speakers included in the study reveals trends similar to those reported by Otheguy and Zentella for their entire corpus. That is, substantial generational differences in pronoun use are observed. Among the recent arrivals, there is a significant difference in the pronoun rates of Caribbean and Mainland speakers: Caribbeans have significantly higher rates of pronoun use ($\underline{M}$ = 37) than Mainlanders ($\underline{M}$ = 18), $t(7) = 3.82$, $p < .007$. Among speakers raised in NYC, Caribbeans have higher rates of pronoun use ($\underline{M}$ = 33) than Mainland speakers ($\underline{M}$= 24), but this difference is not significant, $t(9) = 1.3$, $p = .227$. In other words, while regional origin significantly conditions pronoun use among the study's recent arrivals, it fails to do so among speakers raised in NYC. These trends are illustrated in Figure 4 below.

[4]It is worth noting once more that the methodology illustrated here is agnostic to the segmental status of such fricatives. That is, it is not the goal of this study to assign fricatives to the traditional segmental categories of [s] and [h]. In fact, the data present substantial resistance to such classification.

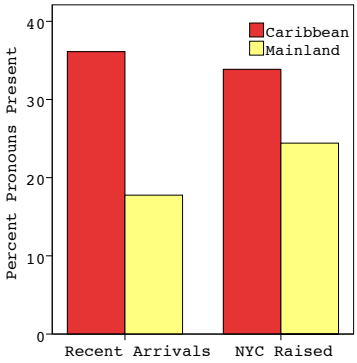

Figure 4: Pronoun use among the current study's speakers.

4.1 Results for /s/ Deletion: the Apparent Generational Homogeneity of Mainlanders

With respect to the presence and absence of frication, several results emerge. First, significant differences in rates of s-deletion are observed along regional lines. Caribbean speakers have a higher rate of deletion than Mainland speakers, $t(3998) = 23.8$, $p < .001$. See Figure 5.

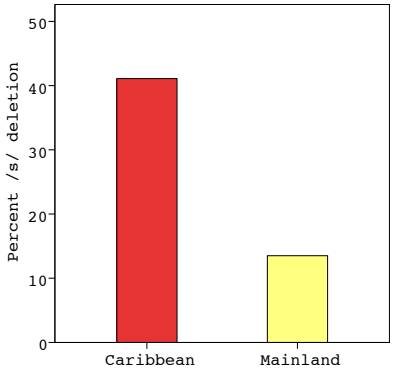

Figure 5: Deletion rates by region.

In parallel with the results for pronouns, the regional difference in deletion rates is diminished in the speech of the NYC raised compared to that of recent arrivals. Consider Figure 6, which presents the deletion rates for each generation, with speakers grouped by region.

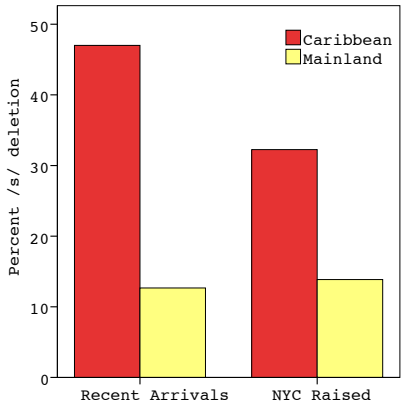

Figure 6: Deletion rates by region and generation.

In contrast to the results for pronouns, where generational differences emerged for both regional groups, there is a key difference in the results for s-deletion. Note in Figure 6 above that the attenuation of regional differences in deletion rates among the NYC Raised is due entirely to a shift in the behavior of Caribbean speakers: NYC raised Caribbeans have a significantly lower rate of s-deletion (M = 37) than their recently arrived counterparts (M = 56), $F = 62.9$, $p < .001$. However, the deletion rates of the two Mainlander groups, 17 percent for the NYC raised compared to 14 percent in the speech of recent arrivals, are very similar, $F = 1.1$, $p < .31$. These results suggest that Mainlanders are relatively stable across generations and that contact has not substantially affected their coda /s/ production.

4.2 Results for Duration and COG: Generational Differences at the Continuous Level

In contrast to the results for deletion, where Mainlanders were relatively homogeneous across generations, results for variability in duration and COG indicate that all NYC raised speakers produce /s/ in a substantially different way than recent arrivals from Latin America. That is, both NYC raised Caribbeans as well as Mainlanders differ significantly from their recently arrived counterparts in terms of the spectrotemporal properties of their /s/ production. In both regional groups, fricative moments are longer in duration and higher in center of gravity in the speech of the NYC raised. Below, Figures 7 and 8 illustrate this finding. Fricatives are plotted by duration and COG for each regional group, with speakers separated by generation.

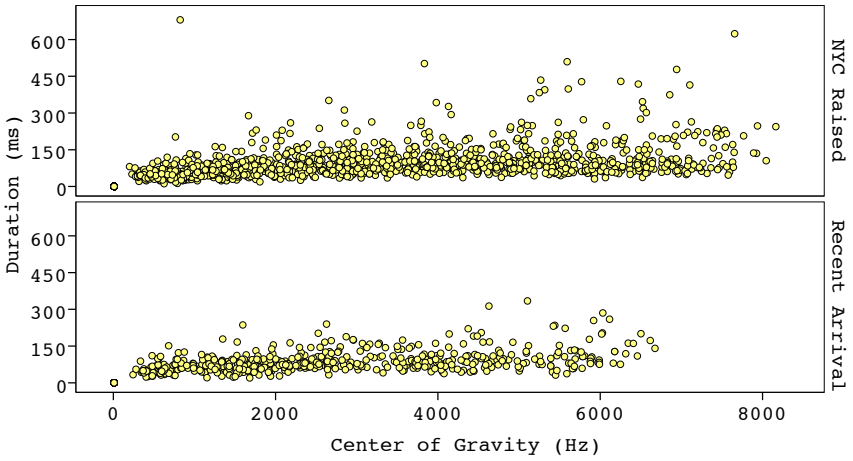

Figure 7: Duration and COG. Mainlanders by generation.

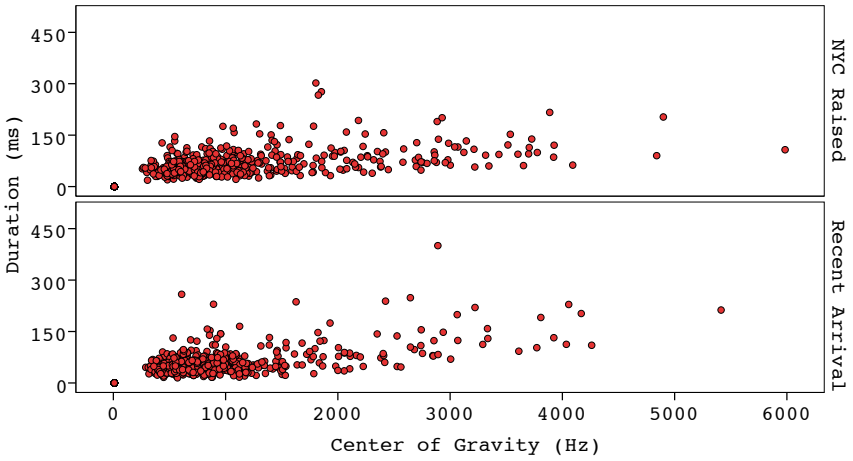

Figure 8: Duration and COG. Caribbeans by generation.

On average, fricatives produced by NYC Mainlanders are 11 milliseconds longer (M = 98 vs. 87) and 700 Hz higher in COG (M = 3400 vs. 2700) than those produced by recent arrivals, $F = 10.8$, $p < .001$, and $F = 37$, $p < .001$, respectively. A similar, if somewhat less robust trend, is observed among Caribbeans: Mean duration is 68 milliseconds in the speech of the NYC raised versus 60 milliseconds for recent arrivals, $F = 12.7$, $p < .001$. Mean COG is 1200 Hz for the NYC raised compared to 1010 Hz for recent arrivals, $F = 15$, $p < .001$, respectively.

To summarize, different patterns are found at discrete and continuous levels of analysis. In the analysis of deletion rates, an overall pattern of dialect leveling emerges, but it is lopsided. NYC raised Caribbeans accommodate towards Mainland norms while the latter group is stable across generations. In contrast, results from the analysis of duration and COG suggest that all NYC raised speakers produce /s/ differently from their recently arrived countrymen. That is, at the level of continuous phonetic substance, Mainlanders are not alike across generations. Among both regional groups, fricatives are longer and higher in COG in the speech of those raised in NYC.

4.3 Linguistic Factors: Logistic and Linear Regression

A similar incongruity between categorical and continuous levels of analysis is observed in the investigation of linguistic factors that condition /s/ production. Below are the results of three regression analyses. The first is a logistic regression. The other two are linear regressions. The dependent variable in the logistic regression is presence vs. absence of frication. The dependent variables in the linear regressions are duration and COG, respectively. In all three analyses, the independent variables are the same: *F1 and F2 of the vowel preceding /s/* (measured at the midpoint of the vowel), *following segment* (consonant, vowel, or pause), *speech rate* (in ms per syllable), *stress* (whether /s/ occurs in a stressed syllable), *lexical frequency* (of the word in which /s/ appears, determined by a frequency dictionary of Spanish), and *morphemic status* of /s/ (whether it is a plural marker, a verbal inflection, or not morphemic).

The logistic regression returns two kinds of information. First, it indicates whether a variable, in the context of the other variables, significantly predicts the presence or absence of frication. Second, it provides a statistic, here a *Wald* value, that ranks the strength of each variable relative to the others. The higher that a *Wald* value is for a given variable, the stronger its role in conditioning pronoun use. Six variables are significant in the analysis. Speech rate has the largest conditioning effect on the presence-absence of frication, followed, in decreasing order of influence, by following segment, morphemic status, lexical frequency, F2 of the preceding vowel, and stress.

Rank	Variable	Wald Value
1	*Speech rate*	159.2**
2	*Following segment*	61.96**
3	*Morphemic Status*	24.5**
4	*Lexical frequency*	20.8**
5	*F2 of preceding V*	19.4**
6	*Stress*	4.6*
7	*F1 of precding V*	2.8 ($p < .09$)

** $p < .001$, N = 4,000, * $p < .05$

Table 3: Logistic Regression–factors constraining the presence/absence of /s/.

There are considerable differences in the linear regression results. First, several factors that significantly condition the presence or absence of frication fail to predict variability in the continuous properties of fricative moments. Consider Table 4 below, in which the relative strength of each independent variable can be determined by comparing the standardized beta coefficients associated with the linear equation that best predicts the value of duration. The larger[5] that the standardized beta is for a given variable, the larger the effect that the factor has on duration. In the regression for duration, only two variables have the same ranking as they did in the logistic regression.

[5]Note that it is the absolute value of the beta coefficients that is relevant to ranking factors. Positivity or negativity of beta values reflects the direction of the factor's effect.

These are speech rate and following segment. In addition to a reshuffling in rankings, three variables that significantly conditioned the presence-absence of frication—namely, morphemic status, stress, and lexical frequency—fail to reach significance as predictors of frication duration.

Rank	Variable	Standardized Beta Coefficients
1	*Speech Rate*	.436**
2	*Following Segment*	-.155**
3	*Preceding V F2*	.100**
4	*Preceding V F1*	.054**
5	*Morphemic Status*	.037 ($p < .051$)
6	*Stress*	.012 ($p < .46$)
7	*Lexical Frequency*	.012 ($p < .47$)

Table 4: Linear Regression–frication duration for /s/.

The results for COG are similar, but not identical, to those for duration. The rank orders of significant variables shift somewhat, with following segment dropping down from second to fourth. In addition, a set variables that made significant contributions to the logistic regression are once again non-significant in the linear regression.

Rank	Variable	Standardized Beta Coefficients
1	*Speech Rate*	.217**
2	*Preceding V F2*	.120**
3	*Preceding V F1*	.097**
4	*Following Segment*	-.086**
5	*Morphemic Status*	-.022 ($p < .33$)
6	*Stress*	-.018 ($p < .36$)
7	*Lexical Frequency*	.001 ($p < .9$)

Table 5: Linear Regression–frication duration for /s/.

Taken together, the logistic and linear regression results suggest that while some linguistic factors globally exert their influence, the predictive potential of others may be restricted to certain domains of expression. For example, speech rate is consistently a very strong factor in predicting both the presence or absence of frication as well as the spectrotemporal properties of fricative moments. Faster speech favors both the deletion of /s/ and also shorter duration and COG when frication is present. By comparison, the predictor following segment, while ranked second in both the logistic regression and in the linear regression for duration, plays a comparatively diminished role in constraining variability in COG. Even more extreme is the behavior of variables like morphemic status, stress, and lexical frequency, which robustly condition s-deletion, but do virtually nothing to account for the continuous properties of fricatives.

5 Conclusion

In the preceding discussion, substantially different pictures of coda /s/ expression emerge at discrete and continuous levels of analysis. The data offer support to the notion that models of sociophonetic variation should take pains to distinguish and relate these levels of variation. In this case, an exclusively continuous approach cannot capture the extent to which regional leveling among Spanish speakers raised in NYC is being lead by Caribbeans accommodating towards the lower deletion rates of the growing Mainlander population. Conversely, an approach that focuses solely on deletion rates overestimates the generational homogeneity of Mainland speakers, who, like their Caribbean counterparts show generational divergence in the spectrotemporal properties of their /s/ production. Furthermore, multivariate analyses of linguistic conditioning factors on /s/

production strongly suggest that the influence of such factors may be restricted to either discrete or continuous levels of expression.

The present study has been largely descriptive in nature, leaving for future research the task of explaining the patterns observed here. Among the questions that remain unanswered, three demand particular attention: (1) Why do deletion rates drop for Caribbeans raised in NYC but remain relatively stable for Mainlanders? (2) Why do all of the Spanish speakers raised in NYC produce fricatives that are longer in duration and higher in COG than recent arrivals? (3) Why do some linguistic factors only condition /s/ production at the discrete or continuous level? Answering these questions will be challenging, but whatever their answers may be, this much seems clear: insofar as our understanding of language variation, contact, and change is built upon the study of variable sound phenomena, we would do well to explore the many dimensions in which speakers produce alternative forms.

References

Bayley, Robert, and Lucinda Pease-Alvarez. 1996. Null and expressed pronoun variation in Mexican-descent children's Spanish. In *Sociolinguistic variation: data, theory, and analysis*, ed. J. Arnold, R. Blake, and B. Davidson, 85–99. Stanford, CA: CSLI.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge University Press.

Cameron, Richard. 1993. Ambiguous agreement, functional compensation, and non-specific *tú* in the Spanish of San Juan, Puerto Rico and Madrid, Spain. *Language Variation and Change,* 5:305–334.

Cedergren, Henrietta J. 1973. The interplay of social and linguistic variables. Doctoral dissertation, Cornell University.

Davies, Mark. 2011. *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. Routledge.

Foulkes, Gerard, and Paul Docherty. 2006. The social life of phonetics and phonology. *Journal of Phonetics*, 34:409–438.

Guitart, Jorge M. 1982. Conservative versus radical dialects in Spanish: implications for language instruction. *Bilingual Education for Hispanic students in the United States*, ed. J.A. Fishman and G.D. Keller, 167–77. New York: Teachers College Press.

Guy, Gregory R. 1980. Variation in the group and the individual: The case of final stop deletion. In *Locating language in time and space,* ed. W. Labov, 1–36. New York: Academic Press.

Labov, William, Paul Cohen, Clarence Robins, and John Lewis. 1968. *A study of the non-standard English of Negro and Puerto Rican speakers in New York City.* Final report, Cooperative Research Project 3288. Vols. I and II.

Labov, William. 1966. *The social stratification of English in New York City*. Cambridge: Cambridge University Press.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1994. *Principles of linguistic change, volume I: Internal factors*. Oxford: Blackwell.

Lipski, John. 1985. /s/ in Central American Spanish. *Hispania*, 68:143–149.

Lipski, John. 1994. *Latin American Spanish*. London: Longman Publishers.

Otheguy, Ricardo, Ana Celia Zentella and David Livert. 2007. Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language*, 83:770–802.

Otheguy, Ricardo, and Ana Celia Zentella. 2012. *Spanish in New York: Language contact, dialect leveling and structural continuity.* Oxford: Oxford University Press.

Scobbie, James, Alice Turk, and Nigel Hewlett. 1999. Morphemes, phonetics and lexical items: The case of the Scottish Vowel Length Rule. *Proceedings of the 14th International Congress of Phonetics Sciences, San Francisco*, 1617–1620.

Silva-Corvalán, Carmen. 1982. Subject expression and placement in Mexican-American Spanish. In *Spanish in the United States: Sociolinguistic aspects*, ed. J. Amastae and L. Elías-Olivares, 93–120. New York: Cambridge University Press.

US Census Bureau. American Community Survey. http://quickfacts.census.gov/qfd/states/36/3651000.html

Watt, Dominic, and Lesley Milroy. 1999. Patterns of variation in Newcastle vowels. In *Urban Voices: Accent Studies in the British Isles*, ed. G. Foulkes and P. Docherty, 25–46. London: Arnold.

Department of Romance Studies
Boston University
Boston, MA 02215
danerker@bu.edu