# The Sociolinguistic Variable: Where is it?

Andrew P. Fleming
*Pennsylvania State University*

# The Sociolinguistic Variable: Where is it?

## Abstract

By examining speakers of Salvadoran heritage in Boston through a model of structural continuity, this paper seeks to understand the how speakers' use of the salient phonological variable of coda /s/ reduction and the supposed non-salient syntactic variable of subject placement change as Spanish speakers spend longer in the United States. Whereas past studies have suggested that Spanish speaker's use of coda /s/ reduction changes according to complex negotiations of sociolinguistic identity, their use of syntactic variables changes due to the effects of the new linguistic environment of the U.S. Rather than addressing these hypotheses directly, the analysis of this study's nine speakers calls attention to the need to better understand the conceptual binaries of salient vs. non-salient and phonological vs. syntactic in the study of sociolinguistic variables. While sociolinguistics often uses terms like coda /s/ reduction and subject placement to refer to "sociolinguistic variables," this study finds evidence that this terminology obscures the nature of salient and non-salient variation among speakers. The data suggests that when Spanish speakers of Salvadoran heritage seek to obscure the regional origins of their speech to avoid raciolinguistic discrimination, they do so by increasing the production of frication of coda /s/ before non-consonants and/or word-finally, and by post-posing more subjects with experiencer-presentative verbs. These results indicate that both these sites of variation in Spanish may hold social meaning in constrained social and linguistic contexts, a finding which demands a new understanding salient and non-salient sociolinguistic variation.

# The Sociolinguistic Variable: Where is it?

Andrew Peter Fleming

## 1 Introduction

In the study of language, the capacity to identify what is variable between speakers stems from the capacity to identify what is not variable. To identify the phonetic distribution and nature of the variable of "R dropping" in English, for instance, the linguist must first understand the shared contexts in which the speakers in question use the "r" sound within a phonological and phonetic grammar, as only then can they discern that one speaker may drop "r" when the prior sound is a vowel and the following sound is not. Without first understanding what unites speakers in their use of "r", there is no grounds to perceive their differences. The continuity of grammars is the means to outline their discontinuities.

Departing from this tenet of variationist sociolinguistics, this paper introduces a new methodology of running regression based on shared behavior between speakers to understand how time spent in the United States correlates with changes in the Spanish of Salvadoran immigrants in Boston. The linguistic variables of syllable-final (coda) /s/ reduction and subject placement are analyzed with this methodology to examine how speakers recently arrived in the United States may use these variables differently from those who have spent longer in the country. While prior research on syllable-final /s/ reduction in U.S. Spanish has focused on the ways in which its speakers use this canonically salient variable to negotiate their sociolinguistic identity in Latinx environments (e.g., Erker 2012, Hernández and Maldonado 2012), research on grammatical subject position has focused on explanations for why Spanish speakers established in the U.S. appear to use more preverbal subjects than those who recently arrived in the country (e.g., Barrera Tobón 2013; Raña Risso 2013; Erker, Ho-Fernández, Otheguy, and Shin 2017).

As a potential way to explain these different trajectories of language change, Erker (2017) suggests that while speakers use coda /s/ reduction to perform sociolinguistic identities in the U.S., the use of presumably low-salient syntactic variables, such as subject placement, changes due to the linguistic convergence of English and Spanish in the minds of bilingual speakers. Whereas the former process is one of agentive language change resulting in a breakdown of regional-group behavior as speakers take on new social identities, the latter is a passive process in which speakers maintain inter-regional differences while experiencing the effects of a more uniform drive of language change. The methodology introduced in this present paper explores this suggestion in Erker 2017 by seeking to identify the precise nature of language change in coda /s/ reduction and subject placement. If coda /s/ reduction is indeed more salient than subject placement, what does this salience manifest as on a structural level? How is it that speakers may use coda /s/ reduction as a salient variable?

The work of Labov and Eckert 2017 suggests that such "style shifting"—i.e. the use of salient variables to perform sociolinguistic identity—occurs on the level of "concrete sounds" in language, meaning that social meaning will more readily to attach to the least abstract levels of linguistic structure, like phonetic sound. While this proposal appears to make sense in the case of canonically salient /s/ reduction, a phonetic trait, the authors' explanation leaves unclear what determines the dichotomy of abstract vs. concrete in language. The authors attempt to explain this distinction through the example of a sound change that results in a phonological merger: the merger, they argue, illustrates an abstract consideration, i.e., one that requires the recognition of one phoneme merging with another, while the accompanying phonetic sound change, they argue, is a more concrete consideration. This explanation, however, raises the question of how the perceiver of the salient sound change interprets the linguistic difference as a deviation from their own phonological and phonetic grammar. In other words, by what means is the phonetic difference isolated and made reproducible as a sound change? In his quantitative study on social salience, Rácz (2013) frames this question in another light by asking, in the case of socially salient linguistic features, what expectations of grammatical systems do they violate?

The broader ambiguity between social meaning and linguistic structure relates back to the ambiguity of the sociolinguistic variable itself. What is it and where is it? In the case of coda /s/ reduction, the matter at first appears straightforward: coda /s/ reduction is a sociolinguistic variable because speakers either produce frication of the /s/ phoneme or not in the syllable-final position. This notion that the variable divides into a binary decision, though, is misleading. Examining Spanish speakers of Caribbean heritage in Boston, Erker and Reffel (in press) demonstrate that the variation of coda /s/ reduction is better accounted for by quantitative measures of frication and duration as opposed to perceptual judgements of deletion, as Spanish speakers generally do not entirely produce or reduce all /s/ frication in the coda position, but rather produce a varying amount of frication according to a set of structural considerations such as phonetic environment, speech rate, size of host word, etc. This gradient nature of /s/ reduction suggests that the conceptualization of the phonological sociolinguistic variable as a choice between two variants (Labov 1972) is more of a methodological consideration rather than empirical description. The question, then, for coda /s/ reduction, is what kinds of fricative reduction make the feature salient? Who gets marked as dropping /s/ and why? Again, by what means is the phonetic difference isolated and made transferable as a so-called "variable"?

These questions concerning the location and nature of the sociolinguistic variable become even more complex and, in a sense, reversed, when applied to the presumed non-salient syntactic variable of subject placement, for which it is less clear whether the label "subject placement" refers to a single sociolinguistic variable, or the result of several sociolinguistic variables converging in one overall rate of preverbal vs. postverbal subjects. While Erker et al. (2017) consider subject placement as a single dependent sociolinguistic variable in a study on Spanish speakers of Cuban heritage in New York City, Raña Risso (2013) focuses subject placement solely with personal pronouns, and Barrera-Tobón (2013) focuses on the subject placement solely in copulative constructions. Adding to this confusion, whereas in coda /s/ reduction the concept of binary variants is a methodological practice to understood a gradient phenomenon, in subject placement the concept of binary variants is an empirical reality understood through a methodological practice of calculating gradient frequencies. Within this model, a specific instance of an overt pronoun or a postverbal subject is not granted the same theoretical potential to index social identity in the way that a coda /s/ token with reduced frication is. Whereas the study of coda /s/ reduction seeks to understand specific tokens, the study of subject placement seeks to understand a general rate of tokens, excluding from consideration the possibility of constructions in which a postverbal subject or overt pronoun can hold sociolinguistic meaning.

To address these concerns over the nature of the sociolinguistic variable, it is necessary to rethink the method of regression analysis used in sociolinguistics. Traditionally, this method consists of selecting independent linguistic variables and independent social variables that account for the variation of the dependent variable of interest. The information for the independent variables is gathered from the dependent variable tokens, and the independent variables are gathered into separate linguistic and social regression models, or are clumped together in one model with both types of independent variables. The speakers are split into subgroups according to one or more social considerations, and the regression models are then carried out separately on the dependent-variable tokens of each subgroup, producing different regression results for each subgroup assessed. These regression results are then organized into hierarchies based on how effectively independent variables and their values predict the dependent variable (resulting in both variable and variable-value hierarchies), and these hierarchies are then compared to see how behavior of the subgroups differ with respect to their use of the dependent linguistic feature in question (Cedergren and Sankoff 1974, Tagliamonte 2012, Guy 2018).

In one of the largest studies ever conducted on U.S. Spanish, Otheguy and Zentella (2012) used this traditional method to analyze subject personal pronoun (SPP) usage among the Spanish speakers of New York City. Collecting data from 140 speakers, the researchers found differences in overt pronoun rates according to speakers' regional heritage and according to how long they have lived in the United Sates. Sorting speakers into groups based on these factors, the researchers found that speakers of Caribbean heritage who have lived all or most of their lives in the U.S. have the highest rate of overt pronoun usage (44%) while speakers of Mainland-Latin-American heritage who have

recently arrived in the U.S. have the lowest (24%). Grouping speakers according to regional heritage and time spent in U.S. as the basis for regression, however, the researchers did not find evidence of wide-spread difference. Rather, they found strong patterns of structural continuity. In short, despite the overall rate differences between subgroups, speakers largely appeared to be guided by the same structural considerations in their use of pronouns. Working with this same Otheguy-Zentella corpus, Raña Risso (2013) and Barrera-Tobón (2013) both find similar evidence of structural continuity in their respective studies on subject placement.

Expanding on these findings, this present study asks that if sociolinguists have found such strong evidence of structural continuity among U.S. Spanish speakers, why then assume structural discontinuity in regression models by dividing speakers into separate groups? Could creating a model that assumes structural continuity among speakers more effectively locate the instances of discontinuity and, consequently, help define what a sociolinguistic variable really is?

## 2 Methodology

### 2.1 Data and Dependent Variables

The study of the present paper examines the speech of nine Spanish speakers of Salvadoran heritage in Boston to see how the use of the variables of coda /s/ reduction and subject placement by speakers who recently arrived to the U.S. differs from those who have spent longer periods of time in the country. The sociolinguistic interviews and coding work for the study are drawn from the Spanish in Boston Corpus (SBC), a collection of interviews of Spanish speakers living in the Greater Boston Area created at Boston University under the direction of Daniel Erker as part of the NSF project, "A Corpus-Based Sociolinguistic Study of Spanish in the Metro-Boston Area" (2014–2018). The coding protocol for each sociolinguistic variable was developed by Erker and a team graduate students at Boston University to examine variation in the data most effectively. The coding work for the study was conducted by a team of researchers at Boston University, including myself, using Praat. While the regression methodology presented in this paper is the result of collaboration between myself and data scientists at Wesleyan University, Connecticut, it is important to stress that the work of the Boston University research team and the support of Professor Erker made this analysis possible.

For the variable of coda /s/ reduction, the segmenting and coding procedures of the present study follow those outlined in Erker (2012) and Erker and Reffel (in press). The /s/ tokens were segmented by researchers according to both the audible perception of frication in the speech stream where /s/ is orthographically represented and according to evidence of frication in the spectrograms and waveforms provided by Praat. The /s/ segments were then measured according to their duration, from the onset of frication to the end, and their average COG value: the latter being a measure of frication 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 (Erker 2012:60). Once the /s/ tokens were segmented and measured according to their phonetic properties, they were then coded for the following linguistic variables: Preceding Vowel; Following Segment; Stress (of the host syllable); Morphological Role; Speech Rate (i.e. syllables of the host word divided by its duration); Number of Segments (in host word); and Universal Lexical Frequency.[1] With respect to the focus on the coda over onset position, this present study acknowledges that /s/ reduction has also been observed in the onset position (Cacoullos and Brown 2002), but, following the findings of Cacoullos and Brown (2003), assumes that /s/ reduction in the onset position is guided by a different set of structural considerations for Spanish speakers than /s/ in the coda position.

For the variable of subject placement, this present study follows the coding procedures of Erker et al. (2017). Per this protocol, tokens of finite verbs with present subjects that could have appeared pre- or post-verbally were analyzed according to the independent linguistic variables of Subject Type, Subject Referent (human or not human), Sentence Type, Clause Type, and Verb Type.

[1]This last independent variable was measured according to Mark Davies' *A Frequency Dictionary of Spanish* (2006).

## 2.2 Regression Methodology Based in Structural Continuity

The methodology developed for this study assumes structural continuity among the speech patterns of its nine participants by placing speaker data in a single group for regression analysis as opposed to relying on speaker subsets. Whereas traditional regression methodology anticipates that different groups of speakers use sociolinguistic variables according to different sets of linguistic considerations, the model proposed here assumes that speakers are largely guided by the same considerations in their use of variables. The purpose of this assumption is to more precisely locate the nature of difference between speakers who have spent large portions of their life in the U.S. compared to speakers who have spent relatively little time in the U.S. While the results of traditional regression often indicate that there are indeed differences among speaker groups, they are often unable to pinpoint where these differences are. For example, the Erker et al. (2017) study on Spanish speakers of Cuban heritage in New York admits that while their regression results indicate the existence of tightly-constrained differences between recently-arrived and established speakers, there is a large measure of chance in the variable and variable-value hierarchies resulting from their regression analyses, making it difficult to say what exactly these differences are.

To locate these differences, the present study considers the amount of time speakers have spent in the U.S. as a social variable with two values: newcomer (defined as having spent less than a year in the U.S.) and non-newcomer. This social variable of Life U.S. is paired as an interaction term with each independent linguistic variable in distinct iterations of the model of independent variables. For each dependent variable, the number of mixed-effects regressions run is thus equal to the number of the independent linguistic variables in the model. In the case of subject placement, for example, there are five independent linguistic variables (i.e., Subject Type, Subject Referent, etc.) so five different mixed-effects regression models were run each with the social variable of Life U.S. paired with a different independent linguistic variable. The mixed-effects model used in this study is the same one used in the Rbrul program (Johnson 2009), which relies on the lmer4 package in RStudio (Bates, Mächler, Bolker, and Walker 2015).[2]

The results of these regressions with interaction terms are then analyzed using the Visreg package (Breheny and Burchett 2017) in RStudio, which graphically displays the results of regression in predicted proportions. Using the "by" argument of this function, the variable-value hierarchies of the independent linguistic variable paired with the Life U.S. interaction term are displayed side-by-side for newcomer and non-newcomer speaker groups. This method consequently allows for an efficient analysis of how Life U.S. may moderate the behavior of speakers. Notably, while the Visreg function allows for the display of regression results in log odds, the values of predicted means and predicted proportions are used here to allow for a clearer comparison and interpretation of behavioral differences and similarities among speaker groups. In the case of the categorical syntactic variable of subject placement, the predicted proportions indicate the probability that when a variable value occurs, a postverbal subject will occur with it. For coda /s/ reduction, the value indicates the most likely COG and the duration values to occur with the corresponding variable value.

The following exemplifies the R code used to create a mixed-effects model for the variable of subject placement with an interaction term Life U.S. on Verb Type, and includes the code to look at the results of that model using Visreg.

- library(visreg)
- library(glmer)
- position.glmm_VERB <- glmer(postpro ~ SUBJ_TYPE+ VERB_TYPE*USlife + SENTENCE_TYPE + SUBJECT_REFERENT + CLAUSE_TYPE+

[2]The only difference is that is that while Rbrul uses sum contrasts—a method that compares the association of independent variable values and the dependent variable against a determined dependent-variable mean—the models for this study use treatment contrasts—a method that compares independent variable values against a base value selected for each independent variable.

```
    (1 | SPEAKER) +
      (1 | TOKEN),
control=glmerControl(optimizer="optimx",optCtrl=list(method="nlminb")),data = position_sal,
family = "binomial")
    summary(position.glmm_VERB, corr=FALSE)
```

- position.glmm_VERB

- visreg_VERB <- visreg(position.glmm_VERB, "VERB_TYPE", by = "USlife", gg = TRUE, scale="response") +
  labs( y ="Proportion", x="Verb Type") +
  theme(legend.position="none")

- visreg_VERB

## 3 Results: Comparison of Newcomer and Non-Newcomer Speakers Using Visreg

Before applying this methodology with Visreg, it is important to understand how the nine speakers of the study use coda /s/ reduction and subject placement according to their experience in the U.S. as a newcomer or non-newcomer. Table 1 summarizes the data between newcomer and non-newcomer groups.

| **Speaker** | **Coda /s/ Reduction** | | **Subject Placement** |
|---|---|---|---|
| | Average COG (hz) | Average Duration (ms) | Proportion Postverbal |
| **Non-newcomer** N=6 | 2917.305 | 98.475 | 0.121 |
| **Newcomer** N=3 | 1584.372 | 78.277 | 0.249 |

Table 1: Summary of Speaker Variable Use.

These results suggest a complex picture of how time spent in the U.S. correlates with changes in the Spanish of Salvadoran migrants. The differences between newcomer and non-newcomer speakers for coda /s/ reduction suggest the general trend observed in Hernández and Maldonado (2012) that when Spanish-speaking Salvadorans arrive in the U.S., some of them begin to produce more frication in their coda /s/ tokens to avoid raciolinguistic discrimination. As a high-salient variable, coda /s/ reduction acts as a marker of a stigmatized Salvadoran and Central-American identity, which, in the context North-American institutions, can lead to greater legal and economic insecurity. For subject placement, the difference between non-newcomer and newcomers reflects the trend observed in Erker et. al. (2017), Raña Risso (2013) and Barrera-Tobón (2013) that as Spanish speakers spend longer in the U.S., they produce fewer postverbal subjects. For both dependent variables, intergroup differences are large: the mean COG and the proportion postverbal subjects for newcomers are nearly half the respective values for non-newcomers. These differences in the rates of reduction between newcomers and non-newcomers, however, are not statistically significant, which is likely due to the small sample sizes of each group (only three speakers constitute the newcomer group) and intergroup variation (speakers were examined according to their individual rates, not by simply aggregating their token data). Taking these limitations into account, the following analysis should be taken as highly speculative, but in line with the rate-based findings of past research.

The results of analysis using the Visreg methodology reveal a more detailed picture about how the Spanish of newcomers and non-newcomers differs. The process of running several models for each dependent variable with an interaction term Life U.S. on each independent linguistic variable reveals that the Salvadoran Spanish speakers of this study are largely guided by the same structural considerations in their use of coda /s/ reduction and subject placement. By finding what is shared

between newcomer and non-newcomers, the regression methodology also reveals what is different. In the case of coda /s/ reduction, it was observed that the difference between newcomers and non-newcomers is largely constrained to phonetic-environment factors and word position. Figures 1 and 2 display the COG and duration values of coda /s/ tokens predicted by the variables of Following Segment and Word Position when all other independent variables are held equal. The data points shown are weighted according to the respective mixed-effects regression models, and the blue lines mark the averages of the weighted data points, thus representing the COG and duration values predicted by each variable value.

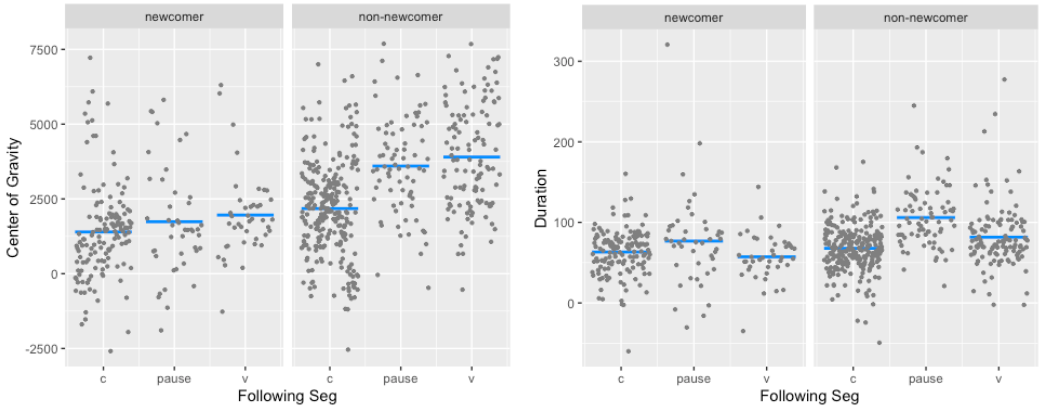


Figure 1: Effect of Following Segment on Coda /s/ COG and Duration According to Life U.S.

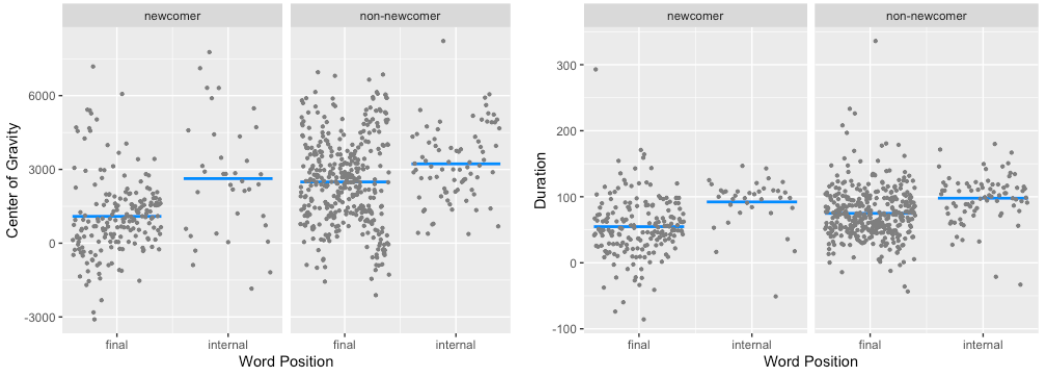


Figure 2: Effect of Word Position on Coda /s/ COG and Duration According to Life U.S.

Figure 1 shows that while non-newcomers tend to produce more coda /s/ frication before pauses and vowels than newcomers, both groups produce relatively similar amounts of frication before consonants. Figure 2 shows that while non-newcomers tend to produce more coda /s/ frication in the word-final position than newcomers, both groups produce relatively similar amounts of frication with coda /s/ tokens in the word-internal position. The observations from these figures are related because all tokens in the word-internal position are inherently followed by a consonant. Within the phonological system of Spanish utilized for this study, a pause cannot appear word-internally and any /s/ that appears before a vowel word-internally moves to the onset position. Thus, the following segment of consonant is the only one that can appear either word internally or finally. If the following segment is a vowel or a pause, then the /s/ token must be in the word-final position.[3] Non-

[3]The phenomenon of coda /s/ shifting to the onset position before a vowel also occurs when coda /s/ is in the word-final position and the following word begins with a vowel. In these cases, the /s/ shifts to join the

newcomer speakers raising /s/ frication before non-consonants consequently may lead to greater frication produced in the word-final position, or, vice versa, greater frication in the word-final position may lead to greater frication appearing to be produced before non-consonants.

For subject placement, the regression methodology reveals a simpler site of specific difference between speakers. Figure 3 illustrates the predicted proportions of postverbal subjects associated with three of types of verbs while holding all other independent variables equal. The blue lines closer to a proportion of 1 indicate a higher likelihood of a postverbal subject while blue lines closer to a proportion of 0 indicate a higher likelihood of a preverbal subject.

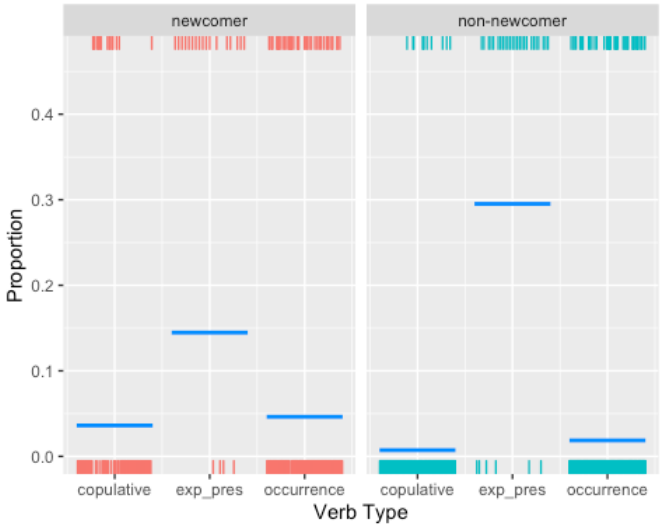


Figure 3: Effect of Verb Type on Subject Placement According to Life U.S.

These results reveal that while for newcomers Verb Type does not predict subject placement (i.e. the differences between the blue lines are not statistically significant), for non-newcomers the verb type of experiencer-presentative predicts a higher rate of postverbal subjects than the verb types of copulative and occurrence. This observed difference between newcomers and non-newcomers suggests that as speakers of Salvadoran heritage spend longer in the U.S., they do not prefer preverbal subjects across all syntactic contexts, but rather adjust their use of subject placement according to the structural consideration of whether the verb is or is not experiencer presentative. Consequently, despite the broader trend towards producing more preverbal subjects, non-newcomer speakers seem to favor postverbal subjects when the verb type is experiencer presentative.

It is important to highlight that, for both coda /s/ reduction and subject placement, the observed differences between variable-value hierarchies are not simply visual. In each case, the interaction term for the observed difference between the newcomer and non-newcomer variable-value hierarchies is statistically significant within its respective regression model. This significance is determined according to whether the variable-value hierarchy of interaction terms (e.g. copulative: non-newcomer; experiencer-presentative: non-newcomer; occurrence: non-newcomer) is different from the variable-value hierarchy without the interaction term (e.g. copulative; experiencer-presentative; occurrence). If the variable-value hierarchy of interaction terms differs from the variable-value hierarchy without interaction terms, then there is evidence of an interaction. In the case of the independent variables for coda /s/ reduction and subject placement not explicitly mentioned in this results section, they were found to confirm that assumption of structural continuity, i.e., the variable-value hierarchies were the same for newcomers and non-newcomers.

---

first-syllable of the following word through a process known as re-syllabification. For the sake of simplicity, this study has chosen to focus on only the lexical level as opposed to the post-lexical level, and consequently does not take this process into account, considering such instances to be tokens of coda /s/ in the word-final position with vowels as the following segment

## 4 Discussion: Simple Complexities

The model of regression introduced in this study based on structural continuity among speakers provides detailed insights into the variable phenomena of coda /s/ reduction and subject placement. In the case of coda /s/ reduction, the results of regression beg the question of whether "coda /s/ reduction" can really be said to be a salient sociolinguistic variable in the case of the study's nine speakers. The data suggests that Spanish speakers of Salvadoran heritage in Boston largely produce syllable-final /s/ according to the same set of structural considerations regardless of how long they have spent in the U.S. The analysis of variable-value hierarchies between newcomer and non-newcomer groups reveals that when Salvadoran speakers face discrimination in the U.S., they seek to obscure their stigmatized region of origin not by radically reconfiguring their use of /s/, but by raising coda /s/ frication before non-consonants and/or in the word-final position when other linguistic considerations are held equal. While future research will reveal which of these two structural considerations guide speakers in their sociolinguistic identity performance, the overall implication from this study's data is that when speakers adjust their use of a salient sociolinguistic variable, they do so according to highly-simplified considerations (i.e., word final or not; before a consonant or not).

To return to the proposal of Eckert and Labov 2017, this observed mechanism of accommodation extends the idea that social meaning attaches to the surface rather than abstract level of linguistic structure to suggest that social meaning in language, in its performative and perceptual use, exist as relationship between a site of variation and a simplified dependent variable. To put this idea in the context of the present study, coda /s/ reduction does not exist as a salient sociolinguistic variable. The sociolinguistic variable is rather the relationship between coda /s/ reduction, an abstract site of variance between speakers, and the structural consideration of whether the following segment is a consonant and/or whether the /s/ token is in the word-final position. The similarities of newcomer and non-newcomer behavior suggest that when Salvadoran Spanish speakers seek to perform a regionally unspecific identity, the frication of coda /s/ before consonants and/or in the word-final position does not hold any social meaning.

Surprisingly, the same dynamic of accommodation appears to be true for the supposed non-salient variation of subject placement, suggesting that, for Spanish-speaking migrants of Salvadoran heritage in Boston, this site of variation may contain a salient sociolinguistic variable. The data suggests that when Salvadoran speakers have spent longer in the U.S., they are more likely to post-pose verbal subjects when used with experiencer-presentative verbs such as "me encanta" ("I love [it]") or "me viene" ("works for me"), when all other linguistic considerations are held equal. Thus, while the general trend observed in past studies is for Spanish speakers from Latin America to use more preverbal subjects as they spend longer in the U.S., this study finds that some Spanish speakers may use more postverbal subjects for specific verbal constructions. This finding suggests that the pre-posing of subjects before experiencer-presentative verbs may hold regional-social meaning as it is a site of deviance from more prestigious Spanish varieties where, when all other linguistic considerations are held equal, a postverbal subject is expected. Within this model, speakers of Salvadoran heritage seeking to mask their regional origins produce more frication in specific instances of coda /s/ and produce more post-verbal subjects with experiencer presentative verbs.

The repeated cache, however, of *when other considerations are held equal*, complicates this proposed relationship of structural continuity and the sociolinguistic variable. In the case of subject placement, there are several other structural considerations that may compete with that of experiencer-presentative verb type to predict a preverbal or postverbal subject. If Salvadoran Spanish speakers in Boston can recognize certain instances of experiencer-presentative verbs with pre-posed subjects as salient, then they must be able to recognize when experiencer-presentative verbs with pre-posed subjects are not salient due to the presence of other structural considerations. For example, the mixed-effects regressions of this study's analysis indicate that preverbal subjects are more likely to occur with personal-pronoun subjects; thus, a personal pronoun subject pre-posed with an experiencer-presentative verb may be less likely to be perceived as deviant and, consequently, less likely to be perceived as socially salient, because of the competing linguistic

considerations. For salient sociolinguistic variables to exist, speakers must interpret their shared variable grammars to recognize sites of difference.

Though obscured by the purportedly simple and concrete nature of phonetic sound change, this same dynamic applies to coda /s/ reduction. For example, a fast speech rate is a great predictor that a coda /s/ token will be short and have a low COG measurement. Thus, for coda /s/ reduction to be perceived as salient, an audience must presumably recognize how fast a speaker is speaking, and interpret whether any potential coda /s/ reduction is due to the fast speech rate or if the speaker also reduces frication word-finally and/or before non-consonants (i.e., the presumably salient consideration). If salient sociolinguistic variables are truly recognized as concrete phenomena, as Eckert and Labov (2017) suggest, and not abstract frequencies, then language users must constantly assess their interlocutors use of sociolinguistic variables in relation to shared constraints.

## 5 Conclusion: Next Steps

The use of regression models based in structural continuity offers a means to rethink and expand the methodological and theoretical paradigms used within sociolinguistics. This paper has introduced a way to analyze sociolinguistic corpus data using the Visreg package in R, but this proposal is only a first step in developing methods to better understand the relationship of structural continuity and discontinuity in language variation and change. Some potential problems of this method include the family-wise error rate incurred by running several models for each dependent variable of interest, and the difficulty of determining whether observed differences between speaker groups are generalizable beyond sample data. Both these issues relate to the wider problem of how to determine whether a significant interaction in the data is a reproducible finding or the inevitable result of running many different regression models and hoping for something statistically significant to pop up. Future work using this methodology must consequently incorporate statistical tools that measure the reproducibility of regression results, such as splitting the data into training and test sets. By working to develop this methodology and the theory behind it, sociolinguistics can more effectively use its findings of structural continuity among language users to better understand their discontinuities.

## References


Barrera-Tobón, Carolina. 2013. *Contact-induced Changes in Word Order and Intonation in the Spanish of New York City Bilinguals*. Doctoral dissertation, The Graduate Center of the City University of New York.

Bates, Douglass, Mächler, Martin, Bolker, Ben, and Walker, Steve. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1).

Breheny, Patrick, and Burchett, Woodrow. 2017. Visualization of regression models using Visreg. *The R Journal,* 9(2):56–71.

Brown, Esther L, and Torres Cacoullos, Rena. 2002. Que le vamoh aher?: Taking the syllable out of Spanish /s/ reduction. In *University of Pennsylvania Working Papers in Linguistics,* 8(3):17–31.

Brown, Esther L, and Torres Cacoullos, Rena. 2003. Spanish /s/: A different story from beginning (initial) to end (final). In *A Romance Perspective on Language Knowledge and Use: Selected Papers from the XXXI Linguistic Symposium on Romance Languages (LSRL),* ed. R. Núñez-Cedeño, L. López, and R. Cameron, 21–38. Philadelphia, PA: John Benjamins Publishing Co.

Cedergren, Henrietta J, and Sankoff, David. 1974. Variable rules: Performance as a statistical reflection of competence. *Language* 50(2):333–355.

Davies, Mark. 2006. *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. Abingdon, U.K.: Routledge.

Eckert, Penelope, and Labov, William. 2017. Phonetics, phonology and social meaning. *Journal of Sociolinguistics* 21(4):467–496.

Erker, Daniel, and Madeline, Reffel. In press. A case study of Caribbeans in Boston and NYC: Describing and analyzing variability in Spanish /s/.

Erker, Daniel, Ho-Fernández, Eduardo, Otheguy, Ricardo, and Shin, Naomi. 2017. Continuity and change in Spanish among Cubans in New York: A study of subject placement with finite verbs. In *Spanish Dialectology: Variation, Contact, and Change,* ed. A. Cuza, Washington, D.C.: Georgetown University Press.

Erker, Daniel. 2012. *An Acoustically Based Sociolinguistic Analysis of Variable Coda /s/ Production in the Spanish of New York City*. Doctoral dissertation. New York University.
Erker, Daniel. 2017. Contact, co-Variation, and sociolinguistic salience: What Mister Rogers knows about language change. In *University of Pennsylvania Working Papers in Linguistics,* 23(2):68.
Hernández, Jose Esteban, and Maldonado, Ruben Armando. 2012. Reducción de /s/ final de sílaba entre transmigrantes salvadoreños en el sur de Texas. *Lengua y Migración* 4:25–42.
Guy, Gregory R. 2018. Variable grammars: Competence as a statistical abstraction from performance. Constructing theories from data. *Questioning Theoretical Primitives in Linguistic Inquiry: Papers in Honor of Ricardo Otheguy*, ed. N. L. Shin, D. Erker, and R. Otheguy. Philadelphia, PA: John Benjamins Publishing Company.
Johnson, Daniel Ezra. 2009. Getting off the Goldvarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1):359–383.
Labov, William. 1972. *Sociolinguistic Patterns* (No. 4). Philadelphia, PA: University of Pennsylvania Press.
Otheguy, Ricardo, and Celia Zentella, Ana. 2012. *Spanish in New York: Language Contact, Dialect Leveling and Structural Continuity*. Oxford University Press.
Rácz, Péter. 2013. *Salience in Sociolinguistics: A Quantitative Approach.* De Gruyter Mouton.
Rocío, Raña Risso. 2013. *A Corpus-based Sociolinguistic Study of Subject Pronoun Placement*. Doctoral dissertation, The Graduate Center of the City University of New York.
Tagliamonte, Sali. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Hoboken, NJ: Wiley-Blackwell.

Department of Spanish, Italian, and Portuguese
Pennsylvania State University
State College, PA 16803
*apf34@psu.edu*