

# UC Santa Cruz

## UC Santa Cruz Previously Published Works

### Title

The status of s in Dominican Spanish

### Permalink

<https://escholarship.org/uc/item/1xp5625d>

### Authors

Bullock, Barbara E  
Toribio, Almeida Jacqueline  
Amengual, Mark

### Publication Date

2014-05-01

### DOI

10.1016/j.lingua.2014.01.009

Peer reviewed

## The status of s in Dominican Spanish

Barbara E. Bullock<sup>a,\*</sup>, Almeida Jacqueline Toribio<sup>b</sup>, Mark Amengual<sup>c</sup>



<sup>a</sup> The University of Texas at Austin, Department of French and Italian, 1 University Station B7600, Austin, TX 78712, USA

<sup>b</sup> The University of Texas at Austin, Department of Spanish and Portuguese, 1 University Station B3700, Austin, TX 78712, USA

<sup>c</sup> Furman University, Department of Modern Languages and Literatures, 3300 Pointsett Highway, Greenville, SC 29613, USA

Received 25 September 2012; received in revised form 20 January 2014; accepted 22 January 2014

Available online 4 March 2014

### Abstract

Theoretical linguistic treatments of the intrusive-s of popular Dominican Spanish (yo[s] tuve < yo tuve 'I had') assume the hypothesis that illiterate speakers have reanalyzed their phonologies so that lexical items no longer contain any trace of coda-s. As a consequence, illiterate speakers are said to restore an s into random syllable codas in an attempt to hypercorrect to a more elevated style. Using natural data gathered from sociolinguistic interviews with Dominicans of diverse literacy levels, we demonstrate that the phonological characterization of intrusive-s in the theoretical literature is incorrect and the hypothesis that illiterate speakers lack etymological /s/ is also shown to be flawed. Instead, the results of a quantitative analysis demonstrate that, despite high rates of s-deletion, overtly manifested s usually corresponds to etymological-s. Intrusive-s arises relatively rarely in our corpus and it appears from this data that lexical and intrusive-s might have distinct linguistic distributions and they may differ in what they mark socially.

© 2014 Elsevier B.V. All rights reserved.

**Keywords:** Hypercorrection; Intrusive segments; Phonology; Dominican Spanish; Literacy

### 1. Introduction

Speakers of vernacular Dominican Spanish occasionally manifest a phonologically curious form of hypercorrection, exemplified by the appearance of a non-etymological s. The inserted segment can occur within a word (1a), at a pause (1b), or between words (1c), as attested by the examples below where the intrusive-s appears in straight brackets. This s can appear in words from any grammatical category, functional or lexical.

- (1) a. Ja[s]cqueline < Jacqueline  
una ba[s]tata fri[s]ta < batata frita 'a fried sweet potato'  
con mi cua[s]tro hijas < cuatro 'with my four daughters'
- b. digo yo[s] < yo 'say I'  
no fueron a Loma[s] '(they) didn't go to Loma'  
en país fuera[s] < fuera 'in a foreign country'
- c. ya[s] tenía < ya tenía 'already had'  
golpes en la[s] cara < la cara 'hits in the face'  
lo que encontraba era[s] pluma < era pluma 'what he found was feathers'

\* Corresponding author. Tel.: +1 814 880 2376.

E-mail addresses: [bullock@austin.utexas.edu](mailto:bullock@austin.utexas.edu) (B.E. Bullock), [toribio@austin.utexas.edu](mailto:toribio@austin.utexas.edu) (A.J. Toribio), [markamengual@gmail.com](mailto:markamengual@gmail.com) (M. Amengual).

This process is unusual from a phonological perspective, and Vaux (2008) goes as far as to classify it as “unnatural” because it appears to target a syllable coda rather than an onset and, as a consequence, a syllable with an intrusive-s becomes more marked than one without. In this, the process differs substantially from its much more well-known English counterpart, the intrusive-r that appears in pre-vocalic position across morpheme and word boundaries in non-rhotic varieties of English, e.g., *draw[r]ing*; *I saw[r] it*.<sup>1</sup>

Although intrusive-r has been a frequent topic for phonological theorizing, Hay and Sudbury (2005) empirically demonstrate that the theories that have been developed to account for this phenomenon are incompatible with the available descriptive evidence. As we will demonstrate here, this is also true for the intrusive-s in Dominican Spanish, which has served for decades as a critical example for theorizing about Spanish phonology (Terrell, 1978, 1981, 1986; Núñez Cedeño, 1980, 1988, 1994; Harris, 2002), and, more recently, about hypercorrection in general (see Bradley, 2006; Vaux, 2008) even though adequate descriptive research on intrusive-s is sparse. The goal of the present article is to provide such empirical data and to demonstrate that the sociolinguistic and structural patterning of intrusive-s evades the characterization that it has to date received.

The social factors that potentially motivate intrusive-s have been largely glossed over by phonologists who have attributed s-hypercorrection solely to the speech of illiterate speakers. It is our view that an understanding of the external factors implicated in intrusive-s realization is essential in evaluating the claim that it is the illiterate speaker’s lack of knowledge of where s is distributed in a lexical item that motivates hypercorrection in vernacular Dominican Spanish. It is the primary purpose of this paper to rebut this claim by presenting an informative and accurate description and analysis of this unusual linguistic feature. For a complete picture, we need to understand who uses historical-s, who uses intrusive-s, and how these different types of s pattern relative to one another. Therefore, this study also considers the overt manifestation of non-intrusive-s (or what we will variably call *historical-s*, *etymological-s*, or *lexical-s*) in the speech of individual speakers who vary by literacy.

Before presenting our empirical study and analysis, we provide an introduction to the understudied phenomenon of intrusive-s, followed by an overview and critique of the extant phonological literature on the topic. This literature overwhelmingly tends to analyze Dominican Spanish hypercorrection as a word-level phenomenon, attempting to account for the linguistic conditions that permit an s to be inserted into a syllable coda. Anticipating our analysis somewhat, we will argue that, in reality, intrusive-s is primarily an external sandhi phenomenon. That is, the data will show that intrusive-s occurs most often *between* words rather than *within* a word and in a context distinct from that of its overtly expressed lexical-s counterpart.

## 2. Background

The Spanish spoken in the Dominican Republic is often described as innovative relative to other national varieties. This observation owes, in part, to the allegedly categorical deletion of word-final and pre-consonantal lexical-s by semiliterate speakers (see Núñez Cedeño, 1980; Terrell, 1981, 1986; Alba, 1990, 2004; López Morales, 1990).<sup>2</sup> It is asserted that these same speakers also insert an s, in a process that has traditionally been referred to as *hypercorrection*. In this article, we prefer to use the term *intrusive-s* because, in the Dominican context, the term *hypercorrect-s* is ambiguous, potentially referring to elevated rates of lexical-s use among individual speakers (see section 3.1).<sup>3</sup> Additionally, we wish to draw a parallel between intrusive-s and similar phenomena in English.

By way of example, consider (2), in which a semiliterate girl presents two instances of intrusive-s (i.e., in *mes pone* and *Eriska*) and a number of deletions of lexical /s/, indicated by apostrophes. Note that, at the same time, she produces two examples of faithfully retained historical-s (i.e., in *estudio* and *escuela*).

- (2) Yo e[s]tudio lo que la profe me[s] pone en la e[s]cuela y lo hago. . . . A mí me ponen número' ha'ta el cien. . . . Yo sé e'cribir pero no sé a leer. . . . Mi nombre e' Eri[s]ka María.  
'I study what the teacher gives me and I do it. . . . They give me numbers up to one hundred. . . . I know how to write but not how to read. . . . My name is Erika María.'  
(female, age 6, semiliterate)

This curious mixture of retention, deletion, and insertion is not uncommon among the semiliterate speakers that we have observed. What we cannot know *a priori* is whether this speaker’s deletions and insertions are random or whether they belie her knowledge of how s is distributed in lexical items. If her use of s is random then any use of [s] mapped to its

<sup>1</sup> See Hay and Sudbury (2005) and references cited therein on English intrusive [r], and see Gick (2002) on English intrusive [l], as in *I saw[l] it*.

<sup>2</sup> The term *semiliterate* used herein comprises those who are functionally illiterate as well as those who have no ability to read.

<sup>3</sup> Alba (2004) uses *hypercorrection* to refer to elevated rates of historical-s expression and *ultracorrection* to refer to realizations of the intrusive-s.

etymological position is accidental, as her use of any intrusive-s would be. Alternatively, if she does have knowledge of historical-s in some or in all lexical items then it is possible that she merely chooses to suppress them variably. Finally, it is equally plausible that she is aware that an s does not belong in certain items (for instance, in her given name, “Erika”) but chooses to insert one for reasons as yet unknown. Of these possibilities, only the first (that her s use is random) has ever been entertained in the theoretical linguistic literature as an explanation for s-deletion and intrusion in popular Dominican Spanish. We review this literature in the following sections.

### 3. The hypothesis of phonological restructuring in Dominican Spanish

In a series of oft-cited articles, Tracy Terrell hypothesized that speakers such as the one quoted in (2) do not have final /s/ in their underlying representations. According to Terrell (1979:610), they have “completely restructured lexicons, in which no word ends in /s/.” Furthermore, he proposes that because they lack knowledge of underlying /s/, speakers with little education do not know where it should be produced. Those who produce a coda-s, he posits, must do so via an optional rule that inserts the segment “*casí a/ aza*” ‘almost at random’ (1986:129). This notion of semiliterate Dominicans as ‘s-less’ speakers has been promoted by generative phonologists, and Dominican intrusive-s has served to bolster numerous theoretical claims about Spanish phonology and about processes of hypercorrection in general ever since (Harris, 1983, 2002; Núñez Cedeño, 1980, 1988, 1994; Núñez Cedeño and Morales-Front, 1999; Vaux, 2001, 2002; Vaux and Nevins, 2008; Bradley, 2006). The standard generative assumption is encapsulated in the following citation:

“In certain varieties of Caribbean Spanish, /s/ has been systematically and completely lost in syllable-final position; syllable-initial /s/ is not affected. I will call this variety of Spanish “lost-s.” For example, standard *estúpido* ‘stupid,’ *caspa* ‘dandruff,’ and *dos* ‘two’ are *etúpido*, *capa*, and *do*. Lost-s speakers realize that their dialect is stigmatized and the butt of jokes, so in certain social situations they attempt to speak “high class,” a style called *hablar fisno*. But lost-s speakers aren’t sure where the s’s are in standard dialect, so their “corrections” are essentially random, missing the target as often as not. For example, standard *hipopótamo* ‘hippopotamus’ may come out *hispopótamo*, *hipospótamo*, *hipopóstamo*, *hipopótamos* or even *hispospóstamos*.” (Harris, 2002:97)

In his publications on Dominican s, then, Terrell makes two strong, interconnected claims that we summarize here: (i) semiliterate Dominicans lack coda /s/ in underlying representations, and, as a consequence, (ii) any coda [s] that they realize must be the product of insertion. These proposals have been influential in Spanish phonology, but they merit a more critical consideration than they have been given in most of the extant literature. In the ensuing discussion, we summon counterevidence to these claims from sociolinguistic and descriptive work on Dominican Spanish.

#### 3.1. The notion of Dominicans as ‘lost-s’ speakers

There exists a large body of literature on the linguistic and social factors that are implicated in the variable articulation of coda-s in Spanish (see Cedergren, 1973, 1979; Terrell, 1978, 1979; Poplack, 1979; Hammond, 1981; López Morales, 1983; Lipski, 1984, 1985; Lafford, 1986; Samper Padilla, 1990; Cepeda, 1995; Sayahi, 2005; File-Muriel, 2007; Lynch, 2009). Coda-s is known to undergo a process of weakening or lenition throughout the Spanish-speaking world, with aspiration or deletion documented in southern Spain, the Canary Islands, and throughout Latin America, with the exception of Mexican highlands and the Andean region (see Canfield, 1962; Lipski, 1994). As aptly remarked by File-Muriel (2007:6), weakening is so widespread, that “/s/ distribution is easier to describe geographically in terms of where s-aspiration and deletion are not found.”

Traditionally, the underlying lexical sibilant has been viewed as presenting three (or more) variants: the conservative [s], the weakened [h], and the radical Ø. Researchers have focused primarily on the extent to which this variable correlates with factors such as geographic region, age, level of education, gender, and the like. Thus, Spanish s is normally considered to provide evidence for a clear-cut and widespread case of change in progress, with movement observed in apparent time from the conservative to the radical use of this linguistic variable, i.e., from retention to deletion: s>h>Ø (see Lapesa, 1965; Lipski, 1984). With regard to the sociolect defined as popular Dominican Spanish, however, it is unclear whether there is still change in progress or whether the change is near completion. Aspiration was reportedly the norm in the 1920s and 1930s (Henríquez-Ureña, 1940), and three decades later, deletion was observed to be well advanced, especially among younger speakers, for whom the general rule was “la pérdida completa del fonema” ‘complete loss of the phoneme’ (Jiménez Sabater, 1975:80). Hence, it would appear that the weakening trend that is in progress in many varieties of Spanish had reached its end state in the popular Dominican form. Such diachronic facts would appear to sustain Terrell’s claim that Dominican speakers have no underlying /s/ in coda position.

In his studies, Terrell observed deletion rates between 92 and 98% among the semiliterate speakers he sampled. So, it is no surprise that his theory of a phonological restructuring in this population gained traction among linguists. What was perhaps especially appealing to some linguists about this hypothesis was that the presence of intrusive-s could then be

accounted for as a simple reversal, where a rule of deletion is reinterpreted by speakers with no knowledge of orthography as a rule of epenthesis (Vennemann, 1972). But the appropriation of Terrell's notion of Dominicans as 'lost-s' speakers by generative linguists was motivated by theory-internal considerations rather than by language-external ones. Thus, their work uncritically adopted the view that /s/ weakening had run its course in Dominican Spanish and that there was no variation with regard to s production among these speakers.

However, there is a good deal of evidence that the realization of s is at least partially dependent on social factors across sociolects in the Dominican Republic. In colloquial speech, retention of /s/ is stigmatized, and those who use overt [s], whether etymological or not, are said to "*hablar fisno*" < *hablar fino* 'speak in a refined way' (Núñez Cedeño, 1988). A picaresque example of Dominicans' attitude toward overt usage of syllable-final /s/ was offered by one of our informants, who claimed that in encountering someone who uses elevated rates of [s], the proper rejoinder should be: "*¡Diablos, comistes espaguetis!*" 'Damn! You ate spaghetti!', in which both historical and intrusive-s's are heard.<sup>4</sup> What is amusing to Dominicans about this witticism is that none of these s's would normally be pronounced in casual speech. While the realization of /s/ has overt prestige in higher registers and formal styles, its absence appears to enjoy covert prestige in the vernacular, and speakers who use too many sibilants are perceived to be trying to speak above their station. Even highly literate speakers are careful not to exceed certain limits of coda-s production. For instance, Alba (2004, 2009) reports that Dominican university students perceive those who produce high rates of s as affected, and those newscasters who reach close to 100% retention rates are judged to be lacking in authenticity. It appears, then, that there is something 'un-Dominican' about using too many overt sibilants.

The literature further suggests that s-realization may be associated with male effeminacy. In particular, Alba observes that in television and theatre, effeminate characters are depicted by actors producing their coda sibilants, and his own informants state that a male speaker's retention of coda-s would call into question his masculinity or virility. While no empirical study of the relation between perceived social characteristics and s-realization among Dominicans has yet been conducted, these reports are at least suggestive that overuse of overt final s is a potential sociophonetic marker that indexes an identity that many Dominican males would fear to claim (see Mack, 2009, 2010a, 2010b, 2011 for studies of s variation and gender stereotypes in Puerto Rico).

In addition to the socio-indexical properties of overt s, there is further available evidence that illiterate Dominicans are not 'lost-s speakers' who arbitrarily add coda-s. If such speakers did exist, we should not find that they are able to adjust their rates of s-realization according to different conversational styles. Their linguistic performance should demonstrate an accidental or random realization of coda-s, where each token produced would as likely be intrusive as lexical. But we already have available evidence that this is not the case. López Morales (1990) tested a small cohort of semiliterate Dominicans in two conversational contexts. This group of speakers deleted lexical-s at a rate of 91.8% in spontaneous speech, which diminished to 68.2% in their more careful style. He interprets these results to imply that speakers do not lack underlying /s/. Rather, they are aware that they elide it in casual speech and are capable of restoring it in their more careful styles. More importantly, instances of coda-s realization among these speakers corresponded overwhelmingly to lexical-s: only 1.3% of their sibilant realization was intrusive. Based on these results, López Morales argues against the proposal that popular Dominican Spanish has undergone phonological restructuring.

The fact that /s/ deletion appeared to be the norm in this variety by the 1970s coupled with the observation that elevated rates of retention are stigmatized today suggests that in contemporary Dominican Spanish a salient sociolinguistic marker is the presence of s rather than its absence. Hence, while most studies of /s/ in other Spanish varieties examine the predictors of deletion, for colloquial Dominican, the analytical perspective should be reversed to consider the conditions under which speakers are observed to *not* delete, as this is the socially marked choice. It should be noted that in making this claim, we distinguish a speaker's linguistic performance (i.e., the *choice* of using one variable rather than another) from his/her linguistic knowledge, unlike Terrell, who conflated the absence of coda-s in production with its absence in the lexical representations of semiliterate speakers. Rather than positing that semiliterate Dominicans have restructured their lexical entries to be devoid of syllable-final s, we suggest that they might be adhering to social norms in which s-deletion indexes 'Dominicanness'. We postulate, then, that the reason semiliterate Dominicans delete s is not because their lexical representations are deficient. Instead, we suggest that there is a strong, socio-indexical preference among semiliterate speakers for the null form, which entails that they will delete /s/ more frequently than they will produce it.

<sup>4</sup> The inserted-s's are underlined here: *Diablos* comistes espaguetis. It should be pointed out that the s in *comistes* is a redundant verbal morpheme inserted into the preterit by analogy with other tense/mood paradigms in which the second person, informal singular form is marked by {s}. This regularization is attested in rural sociolects of Spanish worldwide, but none of these other varieties employ intrusive-s outside of this morphological context.

### 3.2. The proposal of Dominican coda-s as random epenthesis

Terrell's second claim, that coda-s among semiliterate speakers always results from random insertion, is taken up and formalized by Núñez Cedeño (1988), who proposes the rule of epenthesis in (3), by which s is inserted in syllable-final position:

- (3) /s/ [sic] epenthesis  
 $\emptyset \rightarrow s / \_ \sigma$   
 (Núñez Cedeño, 1988:324)

According to this rule, nearly every syllable coda would be a potential landing site for intrusive-s:

“... there is a tendency in the speech of some Dominican speakers to insert an /s/ in the rhyme provided some general constraints of syllable structure and prosody are respected. For example, the regular words *abogado* ‘lawyer’ and *bofe* ‘lung’ can be hypercorrected as one of the following (for convenience, syllable division is provided where /s/ is inserted): *as.bogado*, *abos.gado*, *abogas.do*, *aboga.dos* and *bos.fe*, *bo.fes*.” (Núñez Cedeño, 1994:30)

Núñez Cedeño invoked the s-insertion data in an attempt to account for various properties related to Spanish prosodic structure. For instance, he posited that insertion would be prohibited in the following circumstances: (a) when a syllable rhyme is already saturated by a branching coda, e.g. *cuarto* ‘fourth’ > \**cuarsto*; (b) before a trill, which he argues to be a geminate consonant, e.g., *perro* ‘dog’ > \**pesrro*; and (c) in the case that stress falls on the antepenultimate syllable, e.g., *águila* ‘eagle’ > \**águisla*. It should be noted that the latter prohibition follows from the assumption that Spanish is quantity-sensitive and that stress assignment cannot skip over a heavy penultimate syllable (a proposal formulated in Harris, 1983). Núñez Cedeño and Morales-Front (1999) have, in more recent years, acknowledged that it is unlikely that s-intrusion bears any real relevance for the issue of quantity-sensitivity in Spanish since s-insertion would occur postlexically (that is, after the application of any stress rules) (see also Alvord, 2003; Roca, 2005). Importantly, the central idea espoused in (3) is that s is inserted into a coda, an assertion that has been adopted by other phonologists in their subsequent analyses (Harris, 2002; Bradley, 2006). Moreover, a subsidiary claim that has gone unchallenged is that “s-epenthesis only operates at the lexical level and that the word boundary # prevents /s/ from seeing the following segment” (Núñez Cedeño, 1988:326). We will dispute these claims with data presented below.

As it is articulated, the insertion rule in (3) poses an empirical challenge to a quantitative analysis of the intrusive-s phenomenon in Dominican Spanish. The principle of accountability (Labov, 1994) dictates that any observation of a linguistic variable include not only the context in which the variable *does* appear, but also every potential context in which it *could* appear. Since s-insertion is allegedly largely unconstrained, nearly every syllable coda in a corpus would need to be coded for whether or not s is realized. For instance, a token such as *abogado*, invoked above, should be counted four times, once for each syllable, because theoretically each syllable of the word is as liable to host an epenthetic-s as any other, according to Núñez Cedeño. Given such a preponderance of potential contexts for insertion, the frequency of occurrence of intrusive-s would be vanishingly small.

In any case, counter-evidence to Núñez Cedeño's relatively unconstrained coda insertion rule is available from observational studies. Morgan (1998) quantified and analyzed intrusive-s in the speech of a Dominican tour guide who attracted his attention as a prolific producer of hypercorrect-s. His case study of the approximately 200 tokens of intrusive-s in the ten-minute discourse produced by the guide revealed that insertion was conditioned by phonological environment. Specifically, 81% of insertions occurred in an external sandhi environment, primarily when the following word began with a voiceless stop. Word-internal insertion also appeared nearly exclusively before [p, t, k] although at much lower rates than in the sandhi environment. These data suggest that, contrary to the rule formulated by Núñez Cedeño, s-insertion is not indiscriminate. Further confirmation that this is likely the case comes from Bullock and Toribio (2010), who analyze the phonological context of 103 actual tokens of intrusive-s produced by a variety of Dominicans. The results of their analysis mirror those of Morgan, again challenging the phonological characterization of this phenomenon as random coda insertion.

One of the obstacles to a full understanding of intrusive-s has been the lack of careful, accountable data on the phenomenon. As mentioned, the focus of the relevant research to date has been largely theoretical. The hypercorrect-s data presented by Núñez Cedeño and used elsewhere appear to be a mix of intuited examples and decontextualized citations from Terrell's field observations. Thus, there is a discrepancy between the approach to the work on Dominican intrusive-s, which is largely anecdotal, and the exhaustive research program dedicated to etymological coda-s in Spanish, despite the fact that the two should be related if a theory of rule reversal is to have any currency. Studies of coda-s in Spanish have been quantitative and, at times, comprehensive in analyzing the linguistic and social factors that are predictive of s-realization. Unfortunately, and somewhat surprisingly, none of this variationist work has been focused on Dominican Spanish, although this national variety is said to be the most advanced with respect to the deletion of coda-s. If,



as Terrell has suggested, intrusive-s arises because underlying /s/ no longer exists in popular Dominican Spanish, a quantitative analysis of coda-s in this variety is imperative to confirm or disconfirm the putative absence of this segment among semiliterate speakers. Further, the two types of s-realization—etymological and intrusive—must be studied in tandem and analyzed by reference to the same linguistic-internal and -external factors within the same group of speakers.

#### 4. Justification for the present study

In order to test the claims that semiliterate Dominicans are ‘lost-s’ speakers who introduce an s into random syllable codas, we undertake a careful study of the linguistic-internal and -external factors that regulate lexical-s and intrusive-s in popular Dominican Spanish. The studies of López Morales (1990), Morgan (1998), and Bullock and Toribio (2010) discussed above collectively suggest that it is unlikely that semiliterate Dominican speakers have no underlying, syllable-final s, and it is equally unlikely that “their ‘corrections’ are essentially random, missing the target as often as not,” as Harris (2002:97) declares. However, each of these contributing empirical studies is of very small scope and provides only descriptive, rather than inferential, statistics. Thus, none of them can fully counter the widespread assumption that the intrusive-s, for which popular Dominican Spanish has become so well known, is a product of phonological restructuring among semiliterate speakers. López Morales’s study on the effects of conversational style on s-realization does address the issue of the lexical representation of a small group of semiliterate speakers, but he does not address the phonological distribution of intrusive-s. Additionally, since his results are presented in the aggregate, we cannot know whether there are still speakers within his sample who produce an intrusive-s with the same or with greater frequency than they produce lexical-s. If such speakers were to be found, this would lend credence to the idea that there are some Dominicans who do not know where /s/ belongs.

A second, important empirical issue concerns the characterization of the intrusive-s as enjoying a mostly unrestricted coda epenthesis, in which Nuñez Cedeño’s hypercorrect *bosfe* < *bofe* ‘lung’ would be as likely to occur as the attested example *Eriska* < *Erika* cited in (2). If this were true, it would imply that intrusive-s in Dominican Spanish flouts the distributional regularities of historical-s in Spanish, which appears most frequently before [p, t, k] within a word (File-Muriel, 2007). Morgan’s case study provides a wealth of data regarding the distribution of hypercorrect-s within a single speaker, but we are left knowing little about his speaker’s relative frequency of use and distribution of correct lexical /s/. Similarly, Bullock and Toribio (2010) may have provided data from a larger sample of speakers than Morgan did, but their study, too, fails to relate the speakers’ rates and distribution of insertion with the frequency and patterns of overt realization of historical-s. Thus, we do not know how intrusive-s patterns relative to etymological-s.

#### 5. The study

The present work pursues the study of coda-s realization in Dominican Spanish, specifically with an eye toward testing the claims that s-intrusion is due to illiteracy, that semiliters produce as much intrusive as etymological s, and that the linguistic constraints that regulate lexical-s realization are suspended for intrusive-s. In rectifying the shortcomings of previous studies, we undertake a quantitative analysis to examine the relative contributions of linguistic and extra-linguistic factors as predictors of lexical-s and intrusive-s realization.

##### 5.1. Data collection

The data were collected in the Cibao region of the Dominican Republic by the first two authors.<sup>5</sup> Sociolinguistic interviews were conducted with forty Dominicans. With the exception of a single adult male, the participants represent a fairly homogenous group, all having been born and raised in the same rural, agrarian area of the province of Dajabón, within a small radius of approximately five kilometers. The aforementioned male participant presents the same social profile as the others, though he resides in another rural community nearby. The participant pool comprised 10 adults (5 men and 5 women, age 19–67) and 30 children (14 boys and 16 girls, age 6–18), differing in degree of literacy, as depicted in Table 1.

Literacy was determined in the course of the interview, with questions targeting level of education and participants’ ability to read and write. For the adults in this population, a terminal education below grade 4 generally correlates with a self-reported lack of literacy. The adults who reported low levels of education and little ability to read and write were

<sup>5</sup> Unlike the vast majority of previous studies of Dominican Spanish, our research is focused on the speech of rural speakers, who are largely removed from the potentially standardizing influences of a robust education system or the need or opportunities to accommodate to speakers of diverse regions or social classes (see Bullock and Toribio, 2009).

Table 1  
Participant groups.

	Child		Adult	
	Male	Female	Male	Female
Literate	7	9	3	2
Semiliterate	7	7	2	3

classified as semiliterate. As for the children, all were enrolled in school at the time of the interview, but some had been held back repeatedly precisely because of their inability to read. In our experience it is not unusual to find adolescents who are still in the fourth grade. We therefore questioned each child about his or her capacity to read and write. Since some of them curiously indicated that they could write but not read, we administered a short, simple reading passage to classify them as literate or not.<sup>6</sup> In sum, because the age of a school child in this rural community does not correlate well with literacy, and we are primarily concerned with the effect of literacy in this study, we collapsed young children and adolescents into a single group (Children) and divided them by literacy (Literate and Semiliterate).

Data were collected in individual, semi-structured interviews by the first authors as part of an on-going ethnolinguistic study of rural communities in the Dominican Republic. The interviews were conducted by the second author, a native speaker of Spanish and a native of the region, and recorded by the first author. The sessions took place in various informal settings, including private homes and public spaces, and in every case, there were additional auditors or over-hearers (see Bell, 1984). In a sense, participants were performing before a larger audience than is normal in a sociolinguistic interview and this could have caused them to adjust their style upwards. However, and notably, conversations are rarely private in this rural context and it is typically the case that a conversation between two people will attract an audience of over-hearers, ratified or not. The sessions with the adults, which averaged approximately 45 min, elicited casual speech through discussion of personal history and issues surrounding life in the countryside. The sessions with the children focused on themes relating to their studies and were carried out at school, so the interviews were brief due to time constraints (lasting approximately 15 min). Aside from the interview format, we also elicited additional data via a story-retelling task, a methodology that has proven to be particularly useful for us in collecting similar content across participants of diverse levels of literacy. For that purpose, the researcher told various folk tales (adapted from Zevallos, 1997) and participants retold the tales once they were confident that they could do so; children were usually asked to retell only one story. The data were recorded from each participant via a Shure SM10A head-mounted, unidirectional cardioid microphone onto a Marantz digital recorder at 48 kHz sampling rate, with 16 bit quantization. The microphone is designed for unwanted noise rejection and assures that the audio quality of the field recordings is very clear despite the fact that the data were gathered in the field rather than the laboratory.

## 5.2. Data coding

Each token where etymological coda-s could potentially appear, whether realized or not, was coded, with few exceptions. Excluded from the analysis were the lexical items *está* (3rd p.sg. pres.copula) and *usted* (2nd p.sg. formal pronoun), high frequency words that are invariably reduced to 'tá' and 'te', respectively, in colloquial speech. We also excluded words with final sibilants that were followed by words with initial sibilants, as it would prove difficult to ascertain the affiliation of the /s/. And we disregarded a few bare nouns that appeared in syntactic contexts where it was impossible to determine whether the item would take a plural -s morpheme, e.g., *encontró una parcela de habichuela(s)* '(he) found a plot of bean(s)'. As mentioned earlier (section 3.1), s weakening in vernacular Dominican Spanish has nearly reached its end state so that aspiration (as opposed to deletion) is rare. Thus, the coding was done auditorily and the classification of the dependent variable was binary (presence versus absence of s). The auditory analysis was performed jointly by the three researchers. Each token file was opened in Praat's editor window (Boersma and Weenink, 2009), played, and visually inspected. Ambiguity was infrequent and only arose when we were unable to decide whether aspiration (versus silence) was present or not. Of over 3500 tokens coded, less than 20 were ambiguous. These tokens were submitted to a close visual inspection for the presence of high frequency energy in the spectrum. The tokens with visible energy were coded as aspirated ( $n = 17$ ) and collapsed with the s class; those without were discarded. Crucially, we also coded all intrusive-s segments and tagged them as non-lexical to distinguish these from the other realized sibilants, which were

<sup>6</sup> Many of the children appear to equate sounding out words with writing. For instance, the child whose speech was excerpted in (2), when asked to spell her name, sounds it out in syllables: E-ri-ka Ma-rí-a. Recall, however, that she elsewhere pronounced it Eriska.



tagged as lexical. Because the children's interviews were substantially shorter than the adults', we coded the full sample of each child's speech, for an average of approximately 50 tokens per child. And since we coded the children's speech from the beginning of the interview, we also coded the recordings for the adults in the same way, up to approximately 200 tokens for each adult, to roughly balance the token count between the two age groups.

### 5.2.1. *Dependent variable*

Our data were submitted to analyses in which the dependent variable was the retention of lexical-*s* versus its deletion. While we recognize that researchers are increasingly advocating the use of continuous acoustic measurements for the sociolinguistic study of /*s*/ weakening in Spanish (see Erker, 2010; File-Muriel and Brown, 2010), we have chosen to treat the dependent variable of *s*-retention in our study as dichotomous: *s* vs.  $\emptyset$ . In popular Dominican speech, this fricative is overwhelmingly realized with the percept of [s] rather than [h] in coda position. We should note that an audible aspirate often alternates with the sibilant, but in syllable-initial position in rural Dominican Spanish, particularly, in high frequency words like the reflexive pronoun *se* [he], the affirmative *sí* [hi], and in the discourse marker *entonces* [en.ton.he] 'then' (see Bullock and Toribio, 2009). Terrell (1986) notes that coda aspiration is practically non-existent in popular speech, and Alba (1981) similarly remarks that it occurs almost exclusively in the speech of the upper classes, notably the same individuals who also produce the highest rates of overt *s*.

In the present study, none of the 17 fricative realizations that were initially coded as [h] were intrusive segments. These tokens were collapsed with the [s] variable for the analysis because, as argued above, we view the relevant sociolinguistic marker for popular Dominican Spanish as one of segmental presence versus absence. In this respect, we follow Poplack (1980) and López Morales (1990). In doing so, we acknowledge that we are not following the trend of the majority of studies of Spanish *s* weakening, in which [h] is generally collapsed with the null variant. However, given the small number of tokens in our corpus that are at issue, we do not believe that the decision to collapse [h] with [s] is of much consequence in the present case.<sup>7</sup>

### 5.2.2. *Independent variables*

A number of language-internal variables are known to condition *s*-realization in Spanish. These include: its position within a word, the type of following segment, the morphological status and function of *s*, suprasegmental factors such as stress and word length, and the lexical frequency of the token. The literature is vast and points to significant variation among different regional and social varieties of Spanish with respect to this feature. Most relevant to this study is the segmental and syllabic contexts in which an intrusive-*s* appears. Núñez Cedeño and others have hypothesized that an intrusive-*s* can occur in any coda in a word. However, following the extant empirical work of Morgan (1998) and Bullock and Toribio (2010), we hypothesize that intrusive-*s* occurs most often between words and is dependent as well on the following segment. Therefore, the two pertinent linguistic predictors in the present study are position in a word and following segment type.

With respect to position, we coded whether historical or intrusive-*s* occur word internally, between words in what we will refer to as an external sandhi environment, or pre-pausally. Pauses were determined by the presence of silence of approximately 50 ms. To date, there has been little explicit attention to word-internal *s*-realization in Dominican Spanish, but research in other varieties suggests that retention is greater word internally than finally, specifically when *s* is followed by a voiceless stop (see Brown, 2009; Minnick Fox, 2006). The effect of the following segment type on word-final *s* in Dominican Spanish has only been examined in very broad strokes, with attention limited to the contrasts between a following word-initial vowel, word-initial consonant, and a pause. However, as noted, the descriptive work on intrusive-*s* suggests that a significant conditioning factor in the appearance of the epenthetic segment is the presence of a following voiceless stop (Morgan, 1998; Bullock and Toribio, 2010). Therefore, in our analysis, we break down the broad class of consonants by voicing and manner features. In summary, we coded for the following linguistic variables: position in a word (internal/sandhi/pause) and following segment type (vowel (V), voiceless stop (T), voiced approximants (D), liquids (L), nasals (N), and fricatives (F)).<sup>8</sup>

We did not analyze the data for prosodic factors such as stress. Additionally, we did not consider morphological variables such as morpheme status or word class, in part because we analyzed non-functional word-internal *s* as well as word-final *s*, but primarily because these have been shown not to be relevant for Dominican Spanish. Terrell (1981, 1986) and Alba (1990, 2004) considered the role of morphological factors (person/number inflection in verbal and nominal paradigms) on word-final *s*, testing the hypothesis that functional load may favor retention in particular nominal or verbal paradigms. Their results indicated that morphological status is not a factor that contributes to *s*-retention in the speakers sampled.

<sup>7</sup> An anonymous reviewer notes that the categorization of an aspirate with the deletion or the retention class is an empirical question. While we agree, in principle, we had so few [h]-like tokens ( $n = 17$ ) that we collapsed them with the retention class and are unable to reconstruct which they were.

<sup>8</sup> Spanish presents the voiced approximants [β, ð, ɣ] rather than the voiced stops [b, d, g] in the context of preceding /*s*/.

We included lexical frequency as a variable because we believed that it could be potentially relevant as a predictor of lexical-s realization in Dominican Spanish, as it is with other varieties (File-Muriel, 2007, 2009; Brown, 2009; File-Muriel and Brown, 2010). Typically, high frequency lexical items favor phonetic reduction and low frequency items favor conservative pronunciations (Bybee, 2001, 2002; Bybee and Hopper, 2001; *inter alia*).<sup>9</sup> Unfortunately, as yet there is no oral corpus of vernacular Caribbean speech for which frequency counts can be established and sometimes even basic vocabulary can differ markedly from that of the varieties for which oral frequency measures of Spanish exist (e.g., Davies, 2006). Nonetheless, following File-Muriel (2007), we extracted the normalized frequency of each word type in our data from the *Corpus de referencia del español actual* (CREA) and entered it as a continuous independent variable. In CREA, frequencies for inflected forms are listed separately, allowing us to capture quantifiable differences between differently inflected forms of the same lemma. The few words that did not appear in CREA (e.g., *cacata* ‘tarantula’, *bastón* ‘walking stick’) were given a frequency of 0.

We included the macro-social variables of sex (female/male), literacy (literate/semiliterate), style (free conversation/story retelling), and age group (adult/child) in our models. Interestingly, age has not been previously studied for Dominican Spanish, although it is the best indicator of a change in progress in apparent-time studies. But extant studies have been consistent in their findings for the other three variables. The highest rates of retention are found among the hyperliterate classes. For instance, Alba (2004) reports 59% s-retention in the casual speech of Dominicans of upper socioeconomic classes, the most literate portion of the population; this rate diminishes to 10% among the lower classes. In addition, increased retention is found among females across sociolects. In Terrell’s (1986) sample, university-educated women presented 33% retention, whereas university-educated men only 16%. Finally, s-realization is moderated by style, with frequency rates increasing from casual to monitored speech and from semi-structured interviews to defined tasks such as reading passages or word lists. López Morales (1990), for example, found that retention rates among his semiliterate participants increased from 8.2% to 31.8% as the context called for more careful speech. Especially important to the present study is the factor of literacy. In support of his claim that semiliterate Dominicans have restructured their phonologies, Terrell notes only 4% s-retention among this group of speakers.

In addition to the linguistic-internal and external predictors set out above, each participant was entered as a random intercept in the statistical models as was each unique word token (see Johnson, 2009; Tagliamonte and Baayen, 2012). This allowed us to test whether only one or two of the participants were responsible for the tokens of overt s expression or whether there were trends among our groups once the effect of the individual was accounted for. Similarly, the inclusion of the word token as a random term in the model allows us to more accurately gauge the relative contribution of the linguistic internal variables to the retention of s.

### 5.3. Analysis

The data yielded 3498 observations. For the analysis of s-retention, we removed the observations of intrusive-s ( $n = 82$ ). Mixed effects logistic regression models were fitted using Rbrul (Johnson, 2009) in the statistical package R (R Development Core Team, 2008) to examine the factors that are most predictive of the retention of historical-s versus its deletion. For the analysis of historical-s retention, the factors of position (with three levels) and following segment type (with seven levels) were combined into a single ‘super’ factor with twelve levels. We combined these variables because particular positions are incompatible with certain types of following segments and we wished to avoid an unbalanced design with a number of missing, or empty cells.<sup>10</sup> The configuration of the factor that combines position plus the following segment type (pause, voiceless stop, voiced approximant, liquid, nasal, fricative, or vowel) is illustrated in Table 2, where the filled cells indicate the twelve different levels that were used in the regression models. The empty cells denote impossible combinations.

For the models, the fixed linguistic independent variables included the combined factors with the levels shown in Table 2 and word frequency. The social predictors were literacy (literate/semiliterate), sex (female/male), age group (adult/child), and style (free conversation/story retelling). The decision to enter the individual study participants into the model as a random term was motivated by our observation that certain individuals appear to us to produce more intrusive-s than do others. By fitting the data to a mixed effects rather than a fixed effects model, we are more confident that the

<sup>9</sup> The reader is referred to Walker (2012) for a review of the literature that points to a significant role for frequency in processes of phonological reduction as well as for an insightful discussion of findings that challenge that body of work (see also the challenges to word frequency articulated in Brown and Raymond, 2012).

<sup>10</sup> This is akin to the Varbrul knockout problem in which there are no observations in a cell, although in our case, the data were categorically absent because they *could not* be observed (see the empty cells in Table 2). One option would have been to collapse the levels of sandhi and pause position. However, this would still have left us with empty cells for the word-internal case. We also rejected the option of coding a ‘dummy’ observation. We thank Colin Bannard for his counsel on this matter.

Table 2  
The twelve levels of the independent variable that combines position and following segment type.

	Pause (P)	Internal (I)	Sandhi (S)
∅	P∅	–	–
Voiceless stop (T)	–	IT	ST
Voiced approximant (D)	–	ID	SD
Liquid (L)	–	IL	SL
Nasal (N)	–	IN	SN
Fricative (F)	–	IF	SF
Vowel (V)	–	–	SV

Table 3  
Overview of the observations.

Status of s	Literate				Semiliterate				Total
	Child		Adult		Child		Adult		
	M	F	M	F	M	F	M	F	
Intrusion	4	6	2	10	0	11	12	37	82
Retention of lexical-s	114	371	53	151	13	76	49	79	906
Deletion of lexical-s	404	257	371	263	193	191	346	485	2510
Total	522	634	426	424	206	278	407	601	3498

significant trends that we observe are predictive and do not just reflect the behavior of one or two participants in our sample (Johnson, 2009; Drager and Hay, 2012).

When fitting the data to various models, the independent variables of age group, style, and word frequency were not significant predictors either as main effects or in any interaction in the models generated. Below, we present the descriptive statistics on the overall rates of retention, intrusion, and deletion in the data set that includes all the observations (section 6). Next, we show the results of the mixed-effect logistic regression (Johnson, 2009), which analyzes the contribution of gender, literacy, speaker, and the linguistic internal factors to the probability that lexical-s will be retained rather than deleted. In a second analysis, in section 6.2, we describe the distribution of intrusive-s and discuss the correlation between an individual's use of lexical-s versus his or her use of intrusive-s.

## 6. Descriptive statistics

The descriptive statistics shown in Table 3 provide a complete breakdown of observations across groups. The totals, as given, demonstrate high rates of deletion (2510/3498 = 71.7%) among the population as a whole. In contrast, we observe a low incidence of s-intrusion. The reduced number of observations for the semiliterate children, compared to the literate children, is likely due to the fact that their interviews were shorter than the others. As our study is especially concerned with the role of literacy on the realization of an intrusive-s, we collapse the data and display the observations across literate and semiliterate participant groups in Table 4. As shown, literate participants exhibit much greater retention of lexical-s than semiliterate speakers. That is, of all possible tokens, literate speakers retain 34.3% (689/2006) of coda-s's while semiliterate speakers maintain 14.5% (217/1492). But observe that, regardless of literacy, if an [s] is realized, it corresponds more often to historical-s for the semiliterate group as well as for the literate group.

Table 4  
Lexical versus intrusive s among literate and semiliterate speakers.

Status of s	Literate	Semiliterate
Intrusion	22	60
Retention of lexical-s	689	217
Deletion of lexical-s	1295	1215
Total	2006	1492

Table 5  
Rbrul analysis of lexical-s production.<sup>a</sup>

Factors	Log odds	Tokens (N)	Proportion	Centered factor weight
Deviance		2680.161		
df		15		
Grand mean		.265		
<b>Sex</b>				
Female	0.85	1871	0.361	0.701
Male	−0.85	1542	0.149	0.299
<b>Literacy</b>				
Literate	0.971	1982	0.347	0.725
Semiliterate	−0.971	1431	0.152	0.275
<b>Position + following context</b>				
Internal_N	3.804	38	0.500	0.978
Internal_T	1.714	810	0.406	0.847
Sandhi_V	0.569	507	0.323	0.638
Pre-pausal	0.538	829	0.238	0.631
Internal_F	−0.340	10	0.100	0.416
Sandhi_T	−0.695	442	0.208	0.333
Sandhi_D	−0.760	419	0.112	0.319
Sandhi_N	−0.841	216	0.176	0.301
Sandhi_L	−1.133	87	0.126	0.244
Internal_D	−1.171	21	0.048	0.237
Sandhi_F	−1.684	34	0.147	0.157
<b>Token</b>				
Std. dev.	1.411			
<b>Participant</b>				
Std. dev.	1.54			

<sup>a</sup> This model uses centered factor weights following the standards of Goldvarb (Sankoff et al., 2005) sociolinguistic analyses. While this might skew the data toward those factors with more observations, the order of the factors remain the same whether the factor weights are centered or uncentered.

### 6.1. Etymological /s/ retention versus deletion

Table 5 presents the results of the fixed-effects of the regression analysis performed in Rbrul. In this data set, all tokens of intrusive-s were removed and the dependent variable is the retention of historical-s versus its deletion. The best model for the data with speaker as a random effect included the linguistic internal factors ( $p < 0.0001$ ), literacy ( $p < 0.01$ ), and sex ( $p < 0.01$ ) as significant fixed effects. No interactions among the fixed effects in the current model were significant.

As is traditional in variationist analyses, factor weights above .5 are interpreted as favoring s-retention. The following are significant linguistic-external predictors of the appearance of an overt lexical-s: the sex of the speaker (females retain s more than do males) and the literacy of the speaker (literate retain s more than do semiliterates). Four linguistic internal levels of the contextual factor favor the retention of lexical-s: two word-internal environments (before voiceless stops (\_T), nasals (\_N)), one external sandhi context (pre-vocally (\_#V)), and the pre-pausal position. Of the word-internal environments, only the one preceding voiceless stops (e.g., *gu[s]tar* 'to like') has a robust number of tokens ( $n = 810$ ) so the results for s preceding a nasal (e.g. *mi[s]mo* 'same') need to be interpreted with caution.<sup>11</sup>

Overall, an analysis of the fixed linguistic-internal predictors of lexical-s retention indicate that the environment that most highly favors s is the word-internal context before voiceless stops (e.g., *gu[s]tar*) and the sandhi environment before vowels (e.g., *mi[s] hijos* 'my sons'). These findings demonstrate that lexical-s realization is not random but instead partly constrained by linguistic-internal factors.<sup>12</sup> The results of the contribution of the social factors toward predicting s-retention are consistent with those of previous studies of Dominican Spanish (Terrell, 1986; López Morales, 1990) and of

<sup>11</sup> Since there were only 3 observations of word internal-s occurring before a liquid, these tokens were removed from the analysis.

<sup>12</sup> These results are a little unusual, though, in that we would expect pre-pausal position to be a better predictor of retention than pre-consonantal position. Historically, it is generally the case that coda /s/ was retained longer in pre-pausal position. In French, for instance, pre-consonantal s was lost by the 12th or 13th century but pre-pausal s was pronounced up until the 17th century (e.g., *les femmes* [le faməs] 'the women') (Pope, 1934), and varieties of Spanish have shown similar trajectories of change (Terrell and Tranel, 1978).

Table 6  
Counts of intrusive-s by position.

Position	Intrusive s	Lexical s	Deletion	Total tokens
Sandhi _T	36	92	350	478
Sandhi _N	4	38	178	220
Pre-pausal	18	197	632	847
Sandhi _D	3	47	372	422
Sandhi _V	7	164	343	514
Internal _T	12	329	481	822
Sandhi _L	2	11	76	89
Sandhi _F	0	5	29	34
Internal _F	0	1	9	10
Internal _L	0	2	1	3
Internal _N	0	19	19	38
Internal _D	0	1	20	21

s-weakening varieties of Spanish in general (see [File-Muriel, 2007](#) for an excellent overview of this literature). These confirm that women and literate speakers realize higher rates of s than do semiliterate speakers. With regard to the results of the random terms, it should be acknowledged that the standard deviation among the participants is fairly large, indicating that there is a good degree of variability among individuals in this study, even though sex and literacy as main effects contribute significantly toward the variance in the data.

### 6.2. Intrusive-s distribution

The second analysis examines the 82 tokens of intrusive-s that occurred across all speakers. [Table 6](#) presents the cross tabulation for intrusive-s versus lexical /s/ realization in each phonological context. A number of insights can be gleaned from these counts. First, intrusive-s appears to occur more frequently in a sandhi environment when the following word begins with a voiceless stop than in any other context (e.g., *no[s] tenía* '(she/he) didn't have'). It occurs twice as often in this context as it does pre-pausally and it occurs far more often in the pre-consonantal sandhi environment than it does in a similar context before a vowel. Also noteworthy is the fact that, unlike the overt realization of lexical-s, realization of an intrusive-s is relatively rare in word-internal position. When it does occur in this position (only 12 occurrences in the entire data set), it does so only before voiceless stops.<sup>13</sup> Thus, Núñez Cedeño's hypothesized exemplars, *bosfe* and *asbogado/abosgado/abogasdo*, would not be predicted to occur in Dominican Spanish based on this data. The distribution of intrusive-s in this data coincides exactly with the data-driven studies of [Morgan \(1998\)](#) and [Bullock and Toribio \(2010\)](#).

### 6.3. The correlation of lexical and intrusive-s usage

A prediction made by the "lost-s" hypothesis is that there should be an inverse correlation between rates of lexical-s and intrusive-s usage. That is, intrusive-s usage should be higher among those who show lower rates of lexical-s. To test whether or not this hypothesis is supported by our data, we scaled the data using a z-score and conducted a correlation of the relationship between the participants' rates of lexical-s and inserted-s usage. The results of this test showed no significant correlation between rates of s-usage ( $r = .20$ ,  $t = 1.310$ ,  $df = 38$ ,  $p$ -value = 0.198).

It merits pointing out that a manual inspection of the data by individual participant reveals that there are four speakers who produced no s at all during their interviews. All were semiliterate children under the age of ten, two boys and two girls, and each presented between 20 and 30 potential sites in which they could have produced a lexical coda-s. Of our forty participants, only these four could be said to be 'lost-s' speakers. But crucially, they are not responsible for s-intrusion, since they produced no s at all. Additionally, we should note that these four speakers were among the youngest of our participants.

<sup>13</sup> The exhaustive list: a[s]quí 'here', ba[s]tata 'sweet potato', boni[s]to 'cute', caca[s]ta 'tarantula', Eri[s]ka 'Erika', escape[s]ta 'rifle', gali[s]pote 'werewolf', hospi[s]tale 'hospital', ma[s]cuto 'woven carryall', vieji[s]ta 'old woman', diale[s]to 'dialect', and o[s]tavo 'eighth'. We are aware that the s that is expressed in the latter two items may be substituting for an underlying segment—the velar in *dialecto* and *octavo*; what is relevant here is that the substitution occurs before a voiceless stop.

## 7. Discussion

The results of this study demonstrate that semiliterate Dominicans cannot be considered ‘lost-s’ speakers, contrary to what has been proposed in the phonological literature. Although their overall rate of s-retention is low compared to that of literate speakers, of the s’s that they do produce, decidedly more are lexical than intrusive. In fact, no individual speaker in our sample produced as many or more intrusive as historical s. Even the most prolific intrusive-s speaker in our sample produced more lexical-s (61%) than intrusive-s (39%). These collective and individual results presented here argue against the long-held assumption that semiliterate Dominicans have lost /s/ from their underlying representations. This is not to say that particular lexical items have not undergone restructuring for some speakers. For example, the word *fósforo* ‘match’ is so commonly pronounced [‘fo.fo.ro] that it is unlikely that many of our rural participants know that the orthography contains an ‘s’. The word appears eight times in our corpus and not once is the lexical /s/ produced. Even if there were many examples like lexicalized ‘fóforo’, this would not entail that semiliterate speakers had lost coda /s/ entirely from their lexicons, as the ‘lost-s’ hypothesis claims. A strong implication of the present study, then, is that popular Dominican hypercorrection can no longer be motivated by such an assumption.

From the findings of this study, it appears that semiliterate Dominicans do indeed know where historical-s belongs, but that they choose to suppress it at higher rates than do literate speakers. Why would Dominicans choose deletion over retention? Recall that s-realization is arguably ‘un-Dominican’ and potentially evokes negative criticism. Evidence that a choice is involved is also found in the differential rates of lexical /s/ expression across sexes. It is to be remarked that the males sampled, *irrespective of their level of literacy*, strongly disfavor s-retention, expressing only 14.8% of lexical-s. Females, even semiliterate ones, on the other hand, favor retention at higher rates than do males: 18.6% retention among semiliterate females vs. 17.7% retention among *literate* males. These results are not surprising in the Dominican context, where use of overt lexical s is gendered, arguably marking effeminacy among males. The previous phonological literature that assumed that semiliterate speakers had no underlying s failed to take these types of social motivations into account. In sum, what appeared to Terrell and others to signal a lack of knowledge of underlying representations may actually reflect social expectations. The social functions of s use in Dominican Spanish remain to be tested, but our findings are strongly suggestive that the realization versus suppression of lexical-s plays a critical role in identity construction in this population.

Our results also contradict the proposition that s expression among semiliterate speakers, whether lexical or intrusive, can be accounted for by a unique rule of random coda insertion. First, it’s unlikely that one rule would suffice to characterize s-realization because there are crucial differences between lexical-s and intrusive-s distribution. With respect to etymological-s, speakers favor retention word internally before [p, t, k], in a sandhi environment before vowels and, marginally, in pre-pausal position. A strong case for a coda insertion rule, if one can be made at all, would only be potentially supported by insertion in the word-internal context. The pre-pausal position could as easily be phrasal as syllabic, and in the pre-vocalic sandhi environment, the s would generally be interpreted as an onset to the following word (as it is in French liaison) and not as a coda.

With regard to s-insertion, our results demonstrate that its use is favored most in the sandhi environment, but in this case, before [p, t, k]. It occurs very rarely in the environments that are most favorable to lexical-s expression. Its distribution overlaps with that of lexical-s only in pre-pausal position. These findings challenge the characterization of the intrusive-s as the product of a random coda insertion rule. It is constrained by context, hence is not random. But more importantly, the context into which it is inserted is not one that could be unambiguously depicted as a coda position. Specifically, it does not appear to be attracted to a coda as much as it is attracted to a following voiceless stop, as is the word-internal etymological /s/.<sup>14</sup> The attractor in both cases is purely segmental and not prosodic. Thus, the rule of coda insertion, as put forth in the phonological literature, is descriptively inadequate in motivating either lexical-s or intrusive-s expression independently, and it certainly does not describe the two types collectively.

The hypotheses advanced in the phonological literature to motivate intrusive-s among semiliterate speakers crucially rested on the notion that /s/ was absent from lexical representation. This assumption forced theorists to depict all realized coda-s as the product of random insertion, not just the intrusive variety. The problem, however, was that no one had quantitatively addressed the distribution of both types of pronounced sibilants in Dominican Spanish. As we have shown here, the linguistic conditioning environment for the two types of s evades a uniform description.

It may have been the case that it was the saliency of the intrusive-s that attracted the attention of researchers and distracted them from also noticing that semiliterate speakers were producing lexical-s. However, the failure to consider both types of s production together relegated the intrusive-s to the status of linguistic curiosity and removed it from broader

<sup>14</sup> In this, it resembles what Silverman (2003) refers to as a pre-spirant, which in Tarascan, a language isolate spoken in Michoacán, Mexico, alternates with a pre-aspirate before a voiceless stop. It also bears a potential resemblance to the so-called *s-mobile*, where Proto-Indo-European roots alternate between an initial voiceless stop (T) and a \*sT cluster as in Latin *cortex* and Old Church Slavonic *skora* ‘skin’ (Southern, 1999). We are grateful to J.P. Mallory for bringing this fact to our attention.



discussions about sociolinguistic variation and hypercorrection. This is unfortunate because the results of the present study additionally suggest that the intrusive-s may be sociolinguistically distinct from the realized historical-s. Specifically, unlike the overt expression of lexical-s, the use of intrusive-s does not appear to correlate with sex. The same population of males and females who differ significantly in their rates of lexical-s retention do not differ significantly in their rates of s-intrusion.

The failure to fully describe and compare the distribution and function of the two types of s in popular Dominican Spanish has dramatically simplified what is in actuality a phonologically peculiar process. One of the most common explanations for the origin of intrusive segments is rule inversion, where a segment is most likely to be inserted into the environment from which it was most frequently deleted (McCarthy, 1991). In this corpus, 44% of all instances of intrusive-s occur in a sandhi environment before a voiceless stop. However, this is not the environment in which historical-s is most likely to delete in popular Dominican Spanish—it is most likely to delete before fricatives or voiced approximants, contexts in which intrusive-s occurs very rarely. Furthermore, the retention rates of lexical-s in the sandhi environment before [p, t, k] and in pre-pausal position do not differ substantially. Yet rarely is s intrusive at a pause compared to the sandhi context for the speakers sampled here (although see below). These facts are incompatible with a rule inversion explanation of s-intrusion. Alternatively, it is plausible to posit that segments are inserted by analogy to words of similar structure. That is to say, intrusion would occur most where retention is favored. However, in this data, lexical-s retention is highly favored in sandhi environments before vowels (e.g., *mi[s] amigos* ‘my friends’) and in word-internal environments before voiceless stops (e.g., *ha[s]ta* ‘until’), precisely the contexts in which intrusion is infrequent. So, neither rule inversion nor analogy stand as explanations of the origin or development of intrusive-s.

At this juncture, it is instructive to compare intrusive-s with intrusive segments in English. Intrusive-r occurs in varieties that have r-sandhi but that are otherwise non-rhotic. Hay and Sudbury (2005) have demonstrated that for New Zealand English r-intrusion increases as rhoticity decreases. In the English case, [r] is inserted into the same type of environments in which r-sandhi also occurs. The Dominican case is markedly different. Dominicans are not lost-s speakers to the degree that New Zealanders are non-rhotic speakers, and there is no evidence from the present study to suggest that there is a change in progress in Dominican Spanish, since age was not a significant predictor of s-retention. And, as we have demonstrated here, lexical s-sandhi in Dominican Spanish occurs pre-vocally (e.g., *mi[s] amigos*) just as r-sandhi does in the English case. But, to reiterate, this is *not* the favored context for s-intrusion, which occurs pre-consonantly in a sandhi environment (e.g., *ya[s] tenía*). Thus, it is unlikely that a decrease in s expression is responsible for the intrusion that is found in Dominican Spanish. The origin of s-intrusion remains unknown.

## 8. Conclusion

This study has shown that Dominicans do not have deficient lexical representations that lack coda /s/ and that such an assumption cannot account for the phenomenon of s-insertion. Similarly, explanations of intrusive-s based on rule inversion or analogy fail to obtain. We have proposed that semiliterate Dominicans know where s belongs, but choose to express it variably for reasons of identity and covert prestige. Moreover, based on an analysis of linguistic and social predictors, we posit that s-insertion is a process that is phonologically independent of lexical-s retention and it may serve different socio-indexical and communicative functions, though, what these might be remains to be discovered.

The questions posed by the phenomenon of intrusive-s in Dominican Spanish encourage future inquiry. Our study has been restricted to observing the role of a limited number of predictors with regard to the manifestation of intrusive-s in natural speech. Anecdotally, we have noticed that s-intrusion appears prevalently when Dominicans are reading aloud. This suggests that the positive correlation between semiliteracy and intrusive-s in this study may potentially be an epiphenomenon of style (see Foulkes, 1997 on intrusive-r in read speech in Newcastle). Had we performed a reading task with more literate speakers, it could well be that they would have achieved or exceeded the rates of s-intrusion found here among semiliterates. We have also encountered speakers who appear to append an intrusive-s frequently as a redundant phrase-final marker, as in (4):

- (4) a. Como no tenía burro[s], decidí[s] irse a pie[s], como hubiera llovido[s], y había mucho lodo[s].  
 ‘Since he did not have a donkey, he decided to go by foot, since it had rained, and there was a lot of mud.’  
 (male, age 19, literacy unknown)
- b. Había un viejo[s], llamado Lino y había un gallo[s], llamado Quiquiriquí, y abajo de una mata[s], había un lobo.  
 ‘There was an old man, named Lino and he had a rooster, named Quiquiriquí, and under a tree, there was a wolf.’  
 (male, age 13, semiliterate)

None of these speakers are included in the present study, but their productions prompt us to ask whether this is the same form of s-intrusion that we have investigated here or whether the process itself takes variable forms and serves more than one social or discursive function.

Ultimately, this study invites us to ask questions that were precluded by the earlier characterization of intrusive-s that has been offered in the literature. Given that neither inversion nor analogy can explain the development of intrusive-s, we need to further investigate its origin and the reasons why it is geographically restricted to Dominican Spanish, among other s deleting varieties. And since the use of intrusive-s appears not to be gendered, as the use of overt lexical-s is, we must consider its social and communicative function and how it might be evaluated by listeners relative to overt realizations of lexical-s. Additionally, because voiceless stops, not syllable codas, act as ‘attractors’ for intrusive-s across word boundaries, we must examine why s-insertion favors this environment, which clearly violates the phonotactics of Spanish, where word-initial /sC/ clusters have historically been eliminated. It should be evident that much more research must be carried out on this unusual phenomenon and it is our hope that the present work has laid the groundwork for such efforts.

## References

- Alba, O., 1981. Función del acento en el proceso de elisión de la /a/ en la República Dominicana. In: Alba, O. (Ed.), *El español del Caribe: Ponencias del VI simposio de dialectología*. Universidad Católica Madre y Maestra, Santiago, Dominican Republic, pp. 17–26.
- Alba, O., 1990. *Variación Fonética y Diversidad Social en el Español Dominicano de Santiago*. Pontificia Universidad Católica, Santiago, Dominican Republic.
- Alba, O., 2004. *Cómo Hablamos los Dominicanos*. Ramon Jimenés, Santo Domingo, Dominican Republic.
- Alba, O., 2009. *La Identidad Lingüística de los Dominicanos*. Ediciones Librería La Trinitaria, Santo Domingo, Dominican Republic.
- Alvord, S., 2003. The psychology unreality of quantity sensitivity in Spanish: experimental evidence. *Southwest Journal of Linguistics* 20 (2), 1–12.
- Bell, A., 1984. Language style as audience design. *Language in Society* 13, 145–204.
- Boersma, P., Weenink, D., 2009. Praat. Doing phonetics by computer (version 5.104) [Computer program].
- Bradley, T., 2006. Spanish rhotics and Dominican hypercorrect s. *Probus* 18, 1–33.
- Brown, E., 2009. The relative importance of lexical frequency in syllable- and word-final /s/ reduction in Cali, Colombia. In: Collentine, J., García, M.E., Lafford, B., Marcos Marín, F. (Eds.), *Selected Proceedings of the 11th Hispanic Linguistics Symposium*. Cascadilla Proceedings Project, Somerville, MA, pp. 165–175.
- Brown, E., Raymond, W.D., 2012. How discourse context shapes the lexicon: Explaining the distribution of Spanish f-/h- words. *Diachronica* 29.2, 139–161.
- Bullock, B.E., Toribio, A.J., 2009. Correcting the record on Dominican [s]-hypercorrection. In: Colina, S., Olarrea, A., Carvalho, A. (Eds.), *Romance Linguistics 2009*. John Benjamins, New York, pp. 15–24.
- Bullock, B.E., Toribio, A.J., 2010. Reconsidering Dominican Spanish: data from the rural Cibao. *Rivista Internazionale de Lingüística Iberoamericana* 14, 49–73.
- Bybee, J., 2001. *Phonology and Language Use*. Cambridge University Press, Cambridge.
- Bybee, J., 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14, 261–290.
- Bybee, J., Hopper, P., 2001. *Frequency and the Emergence of Linguistic Structure*. John Benjamins, Amsterdam.
- Canfield, D.L., 1962. *La Pronunciación del Español en América*. Instituto Caro y Cuervo, Bogotá, Colombia.
- Cedergren, H., 1973. *The Interplay of Social and Linguistic Factors in Panamá*. (doctoral dissertation) Cornell University.
- Cedergren, H., 1979. En torno a la variación de la s final de sílaba en Panamá: Análisis cuantitativo. In: López Morales, H. (Ed.), *Corrientes actuales en la dialectología del Caribe hispánico*. Editorial Universitaria, Río Piedras, PR, pp. 37–50.
- Cepeda, G., 1995. Retention and deletion of word-final /s/ in Valdivian Spanish (Chile). *Hispanic Linguistics* 6/7, 329–353.
- Real Academia Española: Banco de datos (CREA) [en línea]. Corpus de referencia del español actual. <http://www.rae.es>.
- Davies, M., 2006. *A Frequency Dictionary of Spanish*. Routledge, New York.
- Drager, K., Hay, J., 2012. Exploiting random intercept: two case studies in sociophonetics. *Language Variation and Change* 2, 59–78.
- Erker, D., 2010. A subsegmental approach to coda /s/ weakening in Dominican Spanish. *International Journal of the Sociology of Language* 203, 9–26.
- File-Muriel, R., 2007. *The Role of Lexical Frequency and Phonetic Context in the Weakening of Syllable-final Lexical /s/ in the Spanish of Barranquilla*. (doctoral dissertation) Indiana University, Colombia.
- File-Muriel, R., 2009. The role of lexical frequency in the weakening of syllable-final lexical /s/ in the Spanish of Barranquilla, Colombia. *Hispania* 92, 348–360.
- File-Muriel, R., Brown, E.K., 2010. The gradient nature of /s/-lenition in Caleño Spanish. *University of Pennsylvania Working Papers in Linguistics* 16, 46–55.
- Foulkes, P., 1997. Rule inversion in a British English dialect: a sociolinguistic investigation of [r]-sandhi in Newcastle upon Tyne. *University of Pennsylvania Working Papers in Linguistics* 4, 259–270.
- Gick, B., 2002. The American intrusive l. *American Speech* 72, 167–183.
- Hammond, R., 1981. El fonema /s/ en el español jíbaro: cuestiones teóricas. In: Alba, O. (Ed.), *El español del Caribe: Ponencias del VI simposio de dialectología*. Universidad Católica Madre y Maestra, Santiago, Dominican Republic, pp. 157–169.
- Harris, J., 1983. *Syllable Structure and Stress in Spanish*. MIT Press, Cambridge, MA.
- Harris, J., 2002. Flaps, trills and syllable structure in Spanish. In: Csirmaz, A., Li, Z., Nevins, A., Vaysman, O., Wagner, M. (Eds.), *MIT Working Papers in Linguistics*, vol. 42. MIT Press, Cambridge, MA, pp. 31–108.
- Hay, J., Sudbury, A., 2005. How rhoticity became /r/-sandhi. *Language* 81, 799–823.
- Henríquez-Ureña, P., 1940. *El Español en Santo Domingo*. Biblioteca de Dialectología Hispanoamericana V, Buenos Aires, Argentina.
- Jiménez Sabater, M.A., 1975. *Más Datos sobre el Español de la República Dominicana*. Ediciones Intec, Santo Domingo, Dominican Republic.
- Johnson, D.E., 2009. Getting off the GoldVarb standard: introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistic Compass* 3 (1), 359–383.

- Labov, W., 1994. *Principles of Linguistic Change*, Vol. 1: Internal Factors, Blackwell, Oxford, UK.
- Lafford, B., 1986. Valor diagnóstico-social del uso de ciertas variantes de /s/ en el español de Cartagena, Colombia. In: Núñez Cedeño, R., Páez, I., Guitart, J. (Eds.), *Estudios sobre la fonología del español del Caribe*. La Casa de Bello, Caracas, pp. 53–75.
- Lapesa, R., 1965. *Historia de la Lengua Española*. Las Américas, New York.
- Lipski, J., 1984. On the weakening of /s/ in Latin American Spanish. *Zeitschrift für Dialektologie und Linguistik* 51, 31–43.
- Lipski, J., 1985. /s/ in Central American Spanish. *Hispania* 68, 143–149.
- Lipski, J., 1994. *Latin American Spanish*. Longman, London.
- López Morales, H., 1983. *El Español del Caribe*. Editorial Mapfre, Madrid.
- López Morales, H., 1990. En torno a la /s/ final dominicana: cuestiones teóricas. *Voz y Letra* 1, 129–137.
- Lynch, A., 2009. A sociolinguistic analysis of final /s/ in Miami Cuban Spanish. *Language Sciences* 31, 766–790.
- Mack, S., 2009. *Socially Stratified Phonetic Variation and Perceived Identity in Puerto Rican Spanish*. (doctoral dissertation) University of Minnesota.
- Mack, S., 2010a. A sociophonetic analysis of perception of sexual orientation in Puerto Rican Spanish. *Journal of Laboratory Phonology* 1, 41–63.
- Mack, S., 2010b. Perception and identity: stereotypes of speech and sexual orientation in Puerto Rican Spanish. In: Borgonovo, C., Español-Echevarría, M., Prévost, P. (Eds.), *Selected Proceedings of the 12th Hispanic Linguistics Symposium*. Cascadia Proceedings Project, Somerville, MA, pp. 136–147.
- Mack, S., 2011. A sociophonetic analysis of /s/ variation in Puerto Rican Spanish. In: Ortiz-López, L. (Ed.), *Selected Proceedings of the 13th Hispanic Linguistics Symposium*. Cascadia Proceedings Project, Somerville, MA, pp. 81–93.
- McCarthy, J., 1991. Synchronic rule inversion. In: Sutton, L., Johnson, C., Shields, R. (Eds.), *Berkeley Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*. Berkeley Linguistics Society, Berkeley, CA, pp. 192–207.
- Minnick Fox, M.A., 2006. *Usage-based Effects in Latin American Spanish Syllable-final /s/ Lenition*. (doctoral dissertation) University of Pennsylvania.
- Morgan, T., 1998. The linguistic parameters of /s/-insertion in Dominican Spanish: a case study of qualitative hypercorrection. In: Gutiérrez-Rexach, J., del Valle, J. (Eds.), *Perspectives on Spanish Linguistics*. Ohio State University, Columbus, OH, pp. 79–96.
- Núñez Cedeño, R., 1980. *La Fonología Moderna y el Español de Santo Domingo*. Editora Taller, Santo Domingo, Dominican Republic.
- Núñez Cedeño, R., 1988. Structure-preserving properties of an epenthetic rule in Spanish. In: Birdsong, D., Montreuil, J.P. (Eds.), *Advances in Romance Linguistics*. Foris, Dordrecht, pp. 319–335.
- Núñez Cedeño, R., 1994. The alterability of Spanish geminates and its effects on the Uniform Applicability Condition. *Probus* 6, 23–42.
- Núñez Cedeño, R., Morales-Front, A., 1999. *Fonología Generativa Contemporánea de la Lengua Española*. Georgetown University Press, Washington, DC.
- Pope, M.K., 1934. *From Latin to Modern French with especial consideration of Anglo-Norman*. Manchester University Press, Manchester.
- Poplack, S., 1979. *Function and Process in a Variable Phonology*. (doctoral dissertation) University of Pennsylvania.
- Poplack, S., 1980. Deletion and disambiguation in Puerto Rican Spanish. *Language* 56, 371–385.
- R Development Core Team, 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria ISBN:3-900051-0701. <http://www.R-project.org>.
- Roca, I., 2005. Saturation of parameter settings in Spanish stress. *Phonology* 20 (2), 345–394.
- Samper Padilla, J.A., 1990. *Estudio Sociolingüístico del Español de Las Palmas de Gran Canaria*. Imprenta Pérez Galdós, Las Palmas.
- Sankoff, D., Tagliamonte, S., Smith, E., 2005. *Goldvarb X: A Variable Rule Application for Macintosh and Windows*. Department of Linguistics, The University of Toronto.
- Sayahi, L., 2005. Final /s/ retention and deletion in Spanish: the role of the speaker's type of competence. *Language Sciences* 27, 515–529.
- Silverman, D., 2003. On the rarity of pre-aspirated stops. *Journal of Linguistics* 39, 575–589.
- Southern, M., 1999. *Sub-Grammatical Survival: Indo-European s-mobile and its Regeneration in Germanic*. Journal of Indo-European Studies Monographs, Washington, DC.
- Tagliamonte, S., Baayen, R.H., 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24 (2), 135–178.
- Terrell, T., 1978. Sobre la aspiración y elisión de /s/ implosiva y final en el español de Puerto Rico. *Nueva Revista de Filología Hispánica* 27, 24–38.
- Terrell, T., 1979. Final /s/ in Cuban Spanish. *Hispania* 62, 599–612.
- Terrell, T., 1981. Relexificación en el español dominicano: implicaciones para la educación. In: Alba, O. (Ed.), *El español del Caribe: Ponencias del VI Simposio de Dialectología*. Universidad Católica Madre y Maestra, Santiago, Dominican Republic, pp. 303–318.
- Terrell, T., 1986. La desaparición de /s/ posnuclear a nivel léxico en el habla dominicana. In: Núñez Cedeño, R., Páez Urdaneta, I., Guitart, J. (Eds.), *Estudios sobre la fonología del Caribe*. Ediciones la casa de Bello, Caracas, pp. 117–134.
- Terrell, T., Tranel, B., 1978. Parallelisms between liaison in French and /s/ aspiration and deletion in Caribbean Spanish dialects. *Montreal Working Papers in Linguistics* 10.1, 31–50.
- Vaux, B., 2001. *Consonant Epenthesis and Hypercorrection*. (unpublished manuscript) Harvard University.
- Vaux, B., 2002. *Consonant Epenthesis and the Problem of Unnatural Phonology*. (unpublished manuscript) Harvard University.
- Vaux, B., 2008. Why the phonological component must be serial and rule based. In: Vaux, B., Nevins, A. (Eds.), *Rules, Constraints and Phonological Phenomena*. Oxford University Press, Oxford, UK, pp. 20–60.
- Vaux, B., Nevins, A., 2008. *Rules, Constraints and Phonological Phenomena*. Oxford University Press, Oxford, UK.
- Vennemann, T., 1972. Rule inversion. *Lingua* 29, 209–242.
- Walker, J.A., 2012. Form, function, and frequency in phonological variation. *Language Variation and Change* 24, 397–415.
- Zevallos, A., 1997. *Los Cuentos del Tío Lino*. Lluvia Editores, Lima, Peru.