Intervocalic Voicing and Regressive Voicing Assimilation in L2 Spanish /s/[1]

Chelsea Escalante
The University of California, Davis

Abstract

Although a substantial amount of research has examined possible areas of pronunciation difficulty for L1 English speakers of L2 Spanish, the role that Spanish /s/ may play in the presence of a foreign accent has largely gone unexplored. Because of the distinct status of Spanish and English /s/, prescriptive models of Spanish phonology would predict two possible sources of 'error' in L1 English/L2 Spanish speech: replacing [s] with [z] intervocalically (i.e. *pre[z]idente* 'president') and failing to regressively assimilate /s/ to [z] preceding a voiced consonant (i.e. *mi[s]mo* 'same'). The goal of this preliminary study is to analyze quantitatively the production of /s/ among six L2 speakers, to compare their sibilants to those of native speakers, and to determine the linguistic and extralinguistic factors (following phonological context, orthographic representation, cognate status, proficiency level, word knowledge, and individual difference) that may contribute to the production of non-target /s/. The logistic regression analysis reveals significant differences between the present study's participants as compared to the monolingual Spanish speakers reported in Schmidt & Willis (2011), with the L2 group exhibiting a higher degree of both intervocalic voicing and maintenance of the voiceless variant preceding a voiced consonant. Among the L2s, as experience and proficiency increases, intervocalic voicing decreases while failing to regressively assimilate /s/ increases. Given the nature of regressive voicing assimilation as a gradient and variable process (Schmidt & Willis, 2011), it is suggested that failing to regressively assimilate may not be considered a non-target realization and future studies that explore how a non-assimilated /s/ is perceived are encouraged.

1. Introduction

Recent advances in the field of acoustic analysis have allowed researchers in second language development to analyze students' acquisition of target phonemes in a quantitative manner. Several studies have analyzed areas of particular difficulty for L1 English speakers learning Spanish as an L2. These areas of difficulty include the realization of Spanish vowels (Cobb & Simonet, 2015; Menke & Face, 2010), voiceless stops ([p, t, k]) (Magloire & Green, 1999; Zampini, 1998; Zampini & Green, 2001), voiced approximants ([β, ð, ɣ]) (Díaz-Campos, 2004; Face & Menke 2009; González-Bueno, 1995; Zampini, 1994), rhotics ([ɾ, r]) (Face, 2006; Major, 1986; Reeder, 1998; Rose, 2010; Waltmunson, 2005), the lateral /l/ and palatal nasal /ɲ/ (Díaz-Campos, 2004), as well as

sibilants (Schmidt, 2014).

One area that remains underexplored is the acquisition of Spanish /s/ by English speakers. In Spanish, [s] and [z] are two allophones of one phoneme /s/. The sibilant has historically been described as being realized as [s] in every position except for preceding a voiced consonant, in which case it is realized as [z] (*desde* [dezðe]; *mismo* [mizmo])[2]. This process, called regressive voicing assimilation, occurs both within and between word boundaries. It also traditionally occurs in dialects that maintain coda-position /s/, instead of the 'radical' dialects that aspirate or delete it. In English, assimilation in voicing does occur with sibilants, especially in stem-final obstruent clusters and in suffixes, however, the main direction of the assimilation is progressive rather than regressive (Knight, 2003). In English, /s/ and /z/ are two different phonemes, rendering minimal pairs *bus* [b^{h}ʌs] and *buzz* [b^{h}ʌz], *precedent* [pɹɛsədənt] and *president* [pɹɛzədənt]. In intervocalic position, orthographic 's' can correspond to either the phoneme /s/ or /z/ in English, but in Spanish it is categorically voiceless.

These two important differences between English and Spanish /s/—the fact that voicing assimilation in English is usually progressive rather than regressive and the fact that s-z distinction is phonemic in English but allophonic in Spanish—imply two possible areas of difficulty for L1 English / L2 Spanish speakers. First, L2 speakers may not be able to even perceive regressive voicing assimilation and therefore may be unlikely to produce it themselves. To complicate the issue further, because it is often a variable and gradient process in native speaker speech (Romero, 1999; Schmidt & Willis, 2011), L2 speakers may receive conflicting input regarding assimilation norms that may further inhibit its production. Secondly, L2 speakers may produce [z] in intervocalic position, as they often do in English, yielding *[bizita] (*visita* 'visit') or *[pɹeziðẹ̲nte][3] (*presidente* 'president') with intervocalic voicing.

Despite the differences between English and Spanish /s/, there has been very little empirical scholarship that has explored possible non-target productions of Spanish sibilants (for a descriptive acknowledgement, see Díaz, 2011) and how they may contribute to a perceived foreign accent. To my knowledge, only one study has observed patterns in the production of Spanish /s/ among non-natives (see Schmidt, 2014). In her investigation of the process of regressive voicing assimilation among advanced L2 speakers of Spanish, Schmidt observed that assimilation was not produced by the majority of participants. Of her 14 informants, only four assimilated before voiced consonants and only one of the four did so to a degree similar to native-speaker controls of Mexican Spanish (as reported in Schmidt & Willis, 2011). The study was limited to preconsonantal /s/, however, and did not explore intervocalic /s/. Yet because non-target realizations of both preconsonantal and intervocalic /s/ have the potential to contribute to a foreign accent in L2 Spanish, the current study finds it important to explore the two contexts simultaneously.

This study is a preliminary, multivariate (logistic regression) analysis of /s/ production among beginning, intermediate, and advanced L2 speakers of Spanish. It offers an initial exploration of the linguistic and extralinguistic factors that may contribute to non-target voicing patterns, measuring the weight of such factors as following phonological context, orthographic representation, cognate status, proficiency level, word knowledge, and individual difference. It is designed to address the following research questions:

1. How does /s/ produced by L1 English speakers learning Spanish as an L2 differ from that produced by L1 Spanish speakers?
2. What linguistic and extralinguistic factors constrain the realization of non-target sibilants?

2. Review of the Literature

2.1 Voice Assimilation in Spanish

Quilis (1993) states that in Spanish, /s/, like any

voiceless consonant, may be produced as voiced most of all when it is found in contact with a voiced consonant (p. 251). Assimilation is typically considered a product of gestural overlap (Browman & Goldstein, 1989; Quilis, 1993), whereby the tongue and the glottis begin to change in anticipation for the following consonant before the constriction for the preceding coda position /s/ ends (García, 2013, p. 119). Example (1a) provides a list of those contexts where voicing of /s/ is said to occur, while (1b) lists those contexts in which voicing is not predicted (Hualde, 2005, p. 160):

(1) (a) voiced stops: s.b s.d s.g
nasals: s.m s.n
liquids: s.l s.r
glides: s.y s.w

(1) (b) voiceless stops: s.p s.t s.k
fricatives: s.f s.s s.x
affricate: s.tʃ
vowels: sV.

Schmidt & Willis (2011) explain that there have been a number of reports on coda /s/ voice assimilation that suggest that the process may not be categorical but instead variable (Dykstra, 1955; Obaid, 1973; Torreblanca, 1978, 1986). These accounts, however, according to Schmidt and Willis (2011), are primarily based on impressionistic observations and present conflicting claims regarding the phonetic context in which voiced /s/ may occur, the motivation for voicing, and regional variation (p. 3).

In an effort to explore empirically the variability of regressive voicing assimilation, Schmidt & Willis (2011) use a picture task to elicit oral data from 12 Mexico City participants[4]. Participants produced /s/ in three phonetic contexts: preceding a voiceless consonant (/VsC$_{voiceless}$/), preceding a voiced consonant (/VsC$_{voiced}$/), and in intervocalic position (/VsV/).

Table 2. Tokens collected by Schmidt & Willis (2011) according to following phonetic context[5]

Phonetic Context	*Following Segment*	*Examples*	*N*
/VsC$_{voiceless}$/	/p, t, k/	Óscar, espejos, postres	17
/VsC$_{voiced}$/	/b, l, m, n/	Osvaldo, isla, cisne	15
/VsV/	/a, e, o/	César, narices, payasos	17
Total			49

Upon analysis, Schmidt & Willis (2011) reach several conclusions, the main one being that "voicing assimilation of /s/, although indisputably present, is not a categorical process in Mexican

Table 1. Descriptions and proposals of Spanish /s/ voicing as described by Schmidt & Willis (2011)

Phonetic context of voicing of /s/	Voicing before a voiced consonant only (Schwegler, Kempff, & Ameal-Guerra, 2010)	Cases of voicing before a voiceless consonant, vowel, or pause (Obaid, 1973; Robinson, 1979; Torreblanca 1978)
Status of phenomenon	Phonological rule (Whitely, 2002)	Phonetic Process – gradient and variable (Torreblanca, 1978)
Motivation	Due to syllable-final articulatory relaxation (Torreblanca, 1978)	Due to articulatory gesture overlap (Hualde, 2005; Quilis, 1993)
Regional variation	Occurs only in /s/-conserving varieties (Cotton & Sharp, 1998)	Cases of voiced aspirated-/s/ in /s/-weakening variety (Dykstra, 1955; Jimenez-Sabater, 1975)

Spanish" (p. 18); in the contexts that descriptively predict categorical voicelessness, voicing was present in 5% of the tokens for the /VsC$_{voiced}$/ context and in 9% of the tokens for the /VsV/ context. Perhaps most notable was the fact that for the context that predicts regressive voicing assimilation, /VsC$_{voiced}$/, only 63% of the tokens were voiced. Linguistic and extralinguistic factors were found to account for some of the variation observed in the voicing of /s/ in this variety of Spanish, suggesting that sociolinguistic factors may relate to the voicing process.

Another study that explores regressive voicing assimilation in Spanish is that of Romero (1999). Looking at the process from a gestural phonology perspective, Romero finds that the phonetic patterns observed in his data do not correspond to the traditional phonological account of /s/ assimilation. However, such an account would predict a simple spreading of the laryngeal configurations from a voiced consonant to a preceding /s/, resulting in a single, uninterrupted stretch of laryngeal pulsing in a sequence of /s/ + voiced consonant (p. 1796). Instead, his data suggest a gradual process of assimilation that does not affect all sequences equally. He finds that the magnitude of the laryngeal gesture varies depending on context with single consonants presenting smaller glottal amplitudes than clusters regardless, to a certain extent, of their voicing specifications (p. 1796).

2.2 Voice Assimilation in English

Unlike Spanish, the main direction of /s/ voicing assimilation in English is progressive as opposed to regressive. As Al-Harbi (2005) demonstrates in (2), English stem-final obstruents, with very few exceptions, are voiceless.

(2) Stem-final obstruent clusters with sibilant in English
- (a) Grasp [sp]
- (b) List [st]
- (c) Mask [sk]
- (d) Corpse [ps]
- (e) Spitz [ts]
- (f) Ax [ks]

Suffix voicing patterns in English show patterns of progressive assimilation with /s/. According to Knight (2003), this process is fixed in English as there are hardly any exceptions. Example (3) illustrates how the process functions for plurals, possessives, third person singular present tense verbs, and past tense verbs.

(3) Suffixal voicing pattern

(a) Plurals

ropes	[ps]	robes	[bz]
cats	[ts]	roads	[dz]
lakes	[ks]	rugs	[gz]
giraffes	[fs]	gloves	[vz]

(b) Possessive

Hope's	[ps]	Bob's	[bz]
Robert's	[ts]	Ted's	[dz]
Dick's	[ks]	Greg's	[gz]
Ralph's	[fs]	Steve's	[vz]

(c) Third person singular, present tense

sleeps	[ps]	rubs	[bz]
waits	[ts]	needs	[dz]
talks	[ks]	brags	[gz]
laughs	[fs]	behaves	[vz]

(d) Past tense

stopped	[pt]	robbed	[bd]
walked	[kt]	dragged	[gd]
coughed	[ft]	shaved	[vd]

(Al-Harbi 2005)

In English, the only type of regressive assimilation is found across word boundaries and then only when a voiced word final consonant is followed by a voiceless word initial consonant. It is never the case that a word final voiceless consonant becomes

voiced because of a word initial voiced consonant (Knight, 2003). To account for the distribution of the allomorphs in (3) the strategy in traditional generative phonology has been to posit a single underlying form /z/ for the plural and its exact parallels and to provide a progressive assimilation rule that devoices /z/ after voiceless obstruents.

3. Methodology

Participants were comprised of six students from a public university in California. All are L1 English speakers who have acquired Spanish as an L2 in a primarily classroom setting. Two participants were identified as novices (enrolled in first year, first quarter), two as intermediate (enrolled in Spanish composition), and two as advanced (graduate students in the Spanish department). The participants read 86 words containing /s/ (distributed between /VsC_{voiced}/, /$VsC_{voiceless}$/, and /VsV/ contexts) embedded in a carrier phrase. Within the /VsV/ context, 15 of the /s/ segments were orthographically represented with 'z' whereas 24 were represented with 's'. In all of the stimuli, /s/ was preceded by a vowel to ensure similar phonological environment preceding the sibilant (see Table 3 below).

Recordings were carried out in a sound-proof booth to avoid background noise and submitted to analysis by the acoustic analysis program Praat. Each sibilant was identified aurally and by the visual representation of sibilance: high frequency frication and aperiodicity in the wave form. Voicing was determined by the presence of glottal pulses and regular periodic patterns in the wave form. Each token was measured for total length of the sibilant and then measured for percentage voiced and voiceless.

Each token was then coded for dependent and independent variables. It was determined that two different sets of dependent variables were necessary in order to obtain the information needed to answer the research questions: 1) sibilant production ([s] or [z]) and 2) accuracy (target or non-target). Sibilant production—voicing or voicelessness—was determined by the majority percentage of the token. If the token was more than half voiced, it was labeled as voiced; less than half voicing was labeled as voiceless. The second dependent variable, accuracy, was determined by prescriptive models; /s/ that was realized as [s] in intervocalic position and preceding a voiceless consonant but [z] preceding a voiced consonant was considered target pronunciation while /s/ realized as [z] in intervocalic position or preceding a voiceless consonant or /s/ realized as [s] preceding a voiced consonant was considered non-target pronunciation. Independent variables included following phonetic context, orthographic representation of /s/, individual speaker, speaker level, status of cognate, and knowledge of word. Each factor group is explained in more detail below.

Following phonetic context. Each sibilant was coded for its following phonetic context, that is, if /s/ was followed by a voiced consonant, a voiceless consonant, or a vowel. This was included in order to see if novice-level Spanish learners are sensitive to the general phonetic prescriptive tendencies of native speakers regarding voicing assimilation in front of voiced consonants and a maintenance of voiceless-

Table 3. Stimuli by following phonetic context

Phonetic Context	*Following Segment*	*Examples*	*N*
/VsC_{voiced}/	/b, d, g, l, m, n, r/	béisbol, desde, desgracia, mismo, isla, cisne, Israel	25
/$VsC_{voiceless}$/	/p, t, k/	después, hasta, escuela	22
/VsV/	/a, e, i, o/	Rosa, presentar, visitar, generoso	24
/VzV/	/a, o, u/	plaza, perezoso, azul	15
Total			86

ness before vowels and voiceless consonants.

Orthographic representation of /s/. Each sibilant was coded for its orthography, either 's' or 'z' in order to see if orthography has a role in the production of the sibilant given that /s/ and /z/ are different phonemes in English, with /z/ almost categorically corresponding to [z].

Individual speaker. Each sibilant was coded for individual speaker order to see if there was significant variation between each individual participant.

Level of speaker. Each sibilant was coded for beginning, intermediate, or advanced proficiency level in order to see if length of study and exposure affected /s/ production.

Status of cognate. Each token was coded as either being an English-Spanish cognate or not. Within the cognate group, however, there existed cognates in which the English word's /s/ contained voicing (as in *visitar* 'to visit') and others in which the English word's /s/ was voiceless (as in *disco* 'disk'). Therefore, the cognates were further coded for their English counterpart's [+voice] or [-voice] specificity.

Knowledge of word. Each word in which the token was located was coded for being either known to the participant or unknown. This was determined with a short questionnaire completed by the participant after the recording was completed.

A total of 511 tokens were gathered, coded, and submitted to multivariate software Rbrul (Johnson, 2009) for logistic regression analysis twice—once for each set of dependent variables. Logistic regression analysis fits the research goals of the present study because it allows the researcher to measure the relationship between the binary categorical dependent variable (voiced/voiceless or target/non-target) and the various independent variables (as previously mentioned) to explore the weight of multiple factors on the realization of one variant over another. Rbrul allows the researcher to include individual speakers as random effects, in addition to the fixed factors typically found in linguistic research. Including individual speaker as a random effect mitigates against overestimating the effect of social factors (Gorman & Johnson, 2013). Rbrul results can be interpreted in a similar fashion to data of Goldvarb (Sankoff, Tagliamonte, & Smith, 2012). A factor weight between 0.0 and 0.5 indicates that the factor disfavors use of the variant that has been selected as the application value, with weights closer to 0 indicating a stronger disfavoring effect; a factor weight between 0.5 and 1.0 indicates that the factor favors use of the variant selected as the application value, with weights closer to 1 indicating a stronger favoring effect. However, the results for individual factors must be interpreted in relation to the input probability, or corrected mean, of the whole data set. Thus, a factor weight between 0.5 and 1.0 indicates that the variant defined as the application value is more likely to be used than the overall usage rate in the data set (Bayley & Holland, 2014, p. 394). Finally, Rbrul provides log odds, the natural logarithm for the odds of each factor, percentages, and numbers of tokens for each individual favor. The analysis provides significance levels of each individual factor group through a step-up, step-down procedure, excluding from the model those factor groups that fail to reach significance (Bayley & Holland, 2014, p. 394).

4. Results

Out of the 511 tokens collected, 64 (12.5%) contained voiced sibilants and 447 (87.5%) contained voiceless sibilants. They produced the target sound (according to prescriptive models) in 334 cases (65.4%) and non-target sounds in 177 (34.6%). In order to be able to address the research questions, however, it is necessary to see how the [s] and [z] variants were distributed among the linguistic and extralinguistic factors taken into account. Table 4 and Table 5 (following pages) show the distribution of the total number of tokens by dependent variable, along with their percentages, totals, Rbrul weights, and log odds measurements by factor group.

Table 4. Distribution of tokens by factor group for dependent variable: voiced/voiceless								
Factor group/ factors	N_{voiced}	$\%_{voiced}$	$N_{voiceless}$	$\%_{voiceless}$	*Total*	*%*	*Weight in conditioning voiced*	*Log odds*
Status of cognate p=1.03e-13								
Voiced cognate	45	26.2	127	73.8	172	33.7	0.864	1.852
Voiceless cognate	1	0.7	143	99.3	144	28.2	0.087	-2.353
Non-cognate	18	9.2	177	90.8	195	38.2	0.623	0.500
Speaker level p=0.0133								
Novice	35	20.5	136	79.5	171	33.5	0.764	1.174
Intermediate	25	14.9	143	85.1	168	32.9	0.619	0.485
Advanced	4	2.3	168	97.7	172	33.7	0.16	-1.659
Knowledge of word p=0.0082								
Known	63	13.3	411	86.7	474	92.8	0.743	1.063
Unknown	1	2.7	36	97.3	37	7.2	0.257	-1.063
Individual speaker [random, not tested]								
Heather (novice)	10	11.6	76	88.4	86	16.8	0.396	-0.407
Wendy (novice)	25	29.4	60	70.6	85	16.6	0.601	0.424
Anna (intermediate)	19	22.4	66	77.6	85	16.6	0.620	0.507
Jack (intermediate)	6	7.2	77	92.8	83	16.2	0.379	-0.478
Kylie (advanced)	1	1.2	85	98.8	86	16.8	0.458	-0.151
Evan (advanced)	3	3.5	83	96.5	86	16.8	0.546	0.202
Following phonetic context p=0.323								
VsC_{voiced}	18	12.2	130	87.3	148	29.0	n/s	n/s
$VsC_{voiceless}$	4	3.1	127	96.9	131	25.6	n/s	n/s
VsV and VzV	42	18.1	190	81.9	232	45.4	n/s	n/s
Orthographic representation p=0.0605								
S	45	10.7	377	89.3	422	82.6	n/s	n/s
Z	19	21.3	70	78.7	89	17.4	n/s	n/s
Centered input probability: 0.014 Deviance: 268.959								

Table 5. Distribution of tokens by factor group for dependent variable: target or non-target								
Factor group/factors	N_{target}	$\%_{target}$	$N_{non-target}$	$\%_{non-target}$	*Total*	*%*	*Weight in conditioning non-target*	*Log odds*
Following phonetic context *p*=1.31e-59								
VsC_{voiced}	18	12.2	130	87.3	148	29.0	0.955	3.045
$VsC_{voiceless}$	128	97.7	3	2.3	131	25.6	0.096	-2.240
VsV and VzV	188	81	44	19.0	232	45.4	0.309	-0.805
Status of cognate *p*=5.26e-05								
Voiced cognate	84	48.8	88	51.2	172	33.7	0.717	0.927
Voiceless cognate	121	84.0	23	16.0	144	28.2	0.325	-0.730
Non-cognate	129	66.2	66	33.8	195	38.2	0.451	-0.197
Individual speaker [random, not tested]								
Heather (novice)	54	62.8	32	37.2	86	16.8	0.478	0.090
Wendy (novice)	57	67.1	28	32.9	85	16.6	0.516	-0.064
Anna (intermediate)	48	56.5	37	43.5	85	16.6	0.427	0.296
Jack (intermediate)	57	68.7	26	31.3	83	16.2	0.52	-0.08
Kylie (advanced)	60	69.8	26	30.2	86	16.8	0.54	-0.159
Evan (advanced)	58	67.4	28	32.6	86	16.8	0.519	-0.075
Knowledge of word *p*=0.499								
Known	323	68.1	151	31.9	474	92.8	n/s	n/s
Unknown	11	29.7	26	70.3	37	7.2	n/s	n/s
Orthographic representation *p*=0.406								
S	264	62.6	158	37.4	422	82.6	n/s	n/s
Z	70	78.7	19	21.3	89	17.4	n/s	n/s
Speaker level *p*=0.242								
Novice	111	64.9	60	35.1	171	33.5	n/s	n/s
Intermediate	105	62.5	63	37.5	168	32.9	n/s	n/s
Advanced	118	68.6	54	31.4	172	33.7	n/s	n/s
Centered input probability: 0.257 Deviance: 338.527								

Following phonetic context. Taking the participant group as a whole, this sample of L2 speakers do not seem to be sensitive to assimilation norms in Spanish. For contexts in which /s/ is followed by a voiced consonant—the context in which we would expect the most voicing—only 12.2% of tokens are voiced, suggesting that participants act distinctly from the native speakers of Schmidt &Willis (2011) who voiced at a rate of 63% but similarly to Schmidt's (2014) advanced L2 participants who also showed very low levels of regressive voicing assimilation. Breaking the participants into proficiency level (Table 6), we see a strong tendency for all groups to employ the voiceless variant, even in the /VsC_{voiced}/ context, with advanced speakers assimilating the least of all of the groups.

The participant group as a whole incorporates regressive voicing assimilation roughly one-fifth of the amount that native-speaking counterparts do. Although many tokens contained a voiced portion and a voiceless portion, very rarely did the voicing occur on the right side of the sibilant—where it would be expected in the case of regressive assimilation. Instead, if there was a blend of voicing and voicelessness, the voicing almost always occurred on the left hand side, suggesting that the assimilation is progressive in nature, coming from the voiced vowels that precede the sibilant. Out of the total 511 tokens collected, 288 (56%) were fully voiceless and 41 (8%) were fully voiced, meaning that the remaining 182 contained at least a portion of voicing as well as voicelessness. Out of those 182, only 11 contained regressive assimilation. However, this regressive voicing assimilation was not prototypical at all. In 4 cases, it occurred in the anticipated phonetic context (/VsC_{voiced}/) but in 7 cases, the majority, it occurred in contexts in which voicing does not occur among native speakers.

Table 7. Regressive voicing assimilation by following phonetic context

	Words that contained regressive voicing assimilation
VsC_{voiced}	rasgos, isla, disgusto, fantasma
$VsC_{voiceless}$	Óscar
VsV/VzV	piezas, escuchar, rosa, visitar, naturaleza (x2)

Secondly, the voicing that occurred in these 11 tokens was never found to cover the majority of the sibilant; in all cases, less than half of the sibilant was voiced.

The participants also differed from native speaker norms in terms of intervocalic /s/, producing (as a whole group) the voiced variant at 18.1%, a rate twice that of the native speakers of Schmidt & Willis (2011). Looking at proficiency level, inserting the voiced variant in intervocalic position is more prevalent among novice and intermediate speakers and decreases among advanced speakers.

Table 6. Voicing and Following Phonetic Context by Proficiency Level

	Novice				*Intermediate*				*Advanced*			
Following phonetic context	V	VL	Total	Weight	V	VL	Total	Weight	V	VL	Total	Weight
VsC_{voiced}	12	38	50	0.240	6	42	48	0.125	0	50	50	0.000
$VsC_{voiceless}$	3	41	44	0.068	0	43	43	0.000	1	43	44	0.023
VsV and VzV	20	57	77	0.205	19	58	77	0.247	3	75	78	0.038
Total	35	136	171		25	143	168		4	168	172	

Turning to the dependent variable of target/non-target realizations, the data in Table 5 suggest that the /VsC$_{voiced}$/ context shows the most difficulty for Spanish learners in regards to producing native-like sibilants. In this context, only 12.2% of the tokens were produced on-target. The other two contexts proved to be considerably less difficult for L2 speakers; production matched the target in 81% of the /VsV/ and /VzV/ context and in 97.7% for the /VsC$_{voiceless}$/ context.

Lastly, the Rbrul weights for this factor group underscore the above-mentioned patterns. The fact that following phonetic context was not found to be a significant[6] factor in conditioning the use of one variant over another for the dependent variable voiced/voiceless demonstrates that participants' sibilants did not change significantly based on the segment following the sibilant. On the other hand, we do see significant Rbrul weights when the dependent variable is changed to target/non-target. In that case, we see stark polarization, with the /VsV$_{voiced}$/ context strongly favoring the non-target at 0.955, the /VsV$_{voiceless}$/ context strongly disfavoring the non-target at 0.096, and the /VsV/ and /VzV/ contexts moderately disfavoring the non-target at 0.309.

Orthographic representation. With very few exceptions, orthographic 'z' in English corresponds to the realization of the voiced sibilant [z]. Because this is not the case in Spanish, with 'z' mainly being produced as [s] except when it is followed by a voiced consonant (which did not occur in the word list used in the task), it is logical to expect to see high levels of voicing among early Spanish learners when the sibilant is orthographically represented as 'z'. The results in Table 4 demonstrate that voicing is in fact more common when the sibilant is orthographically represented as 'z' (21.3%) as compared to 's' (10.7%). However, the data in Table 5 suggest that participants seem to be quite aware that orthographic 'z' should be pronounced as [s] with 78.7% of /s/ tokens being produced on-target. Although there were differences in /s/ realization between the two orthographic representations when considering percentages, Rbrul did not find this factor group to be a significant predictor of either of the two dependent variables considered, suggesting that orthographic 'z' does not significantly affect voicing nor errors in pronunciation as might be expected.

Individual speaker[7]. As seen in Table 4, the context of individual speaker demonstrates that there exists definite variation in the amount of voicing present among the different participants, with [z] being produced in only 1 token (1.2%) for the most infrequently-voicing participant to voicing in 25 tokens (29.4%) for the most frequent. The Rbrul weights—which range from 0.396 to 0.620 in conditioning [z]—illustrate that there is not a large degree of polarization among these speakers; however, the factor group is still considered to be a significant predictor[8]. Participants Anna, Wendy, and Evan slightly favor [z] while Heather, Jack, and Kylie slightly favor [s]. Interestingly, there is one person from each level that favors [s] and one from each level that favors [z].

There is even less polarization among individual speakers when it comes to producing the target variant (seen in Table 5), with the most inaccurate speaker producing the target 56.5% of the time and the most accurate 69.8% of the time. Despite this small difference, the factor group was still considered a statistically significant predictor for the variable. Anna and Heather very slightly favor the non-target while Wendy, Jack, Kylie, and Evan slightly favor the target. Since all of the weights hover around 0.5, however, the differences between the speakers are minor.

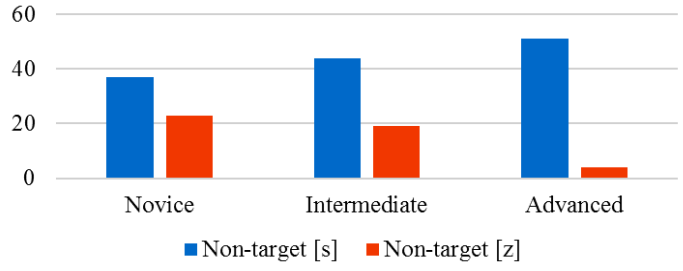
Speaker level. Considering the dependent variable of voicing/voicelessness first, the factor group of proficiency level demonstrates that the groups vary significantly with respect to each other, with the novice group voicing at 20.5%, the intermediate group at 14.9%, and the advanced group at a low 2.3%. The Rbrul weights mirror these tendencies, showing that the novice group highly favors [z] at

0.764, the intermediate group slightly favors the [z] at 0.619, while the advanced group highly disfavors [z] at 0.16.

On the other hand, when the dependent variable target/non-target is considered, as the data in Table 5 illustrate, the rate of on-target variants is quite similar among all three levels. Contrary to expectations, in which it could be assumed that as time of study increases, target sounds would also increase, in actuality, the amount of target sounds remained relatively stable throughout the levels. The novice level hit the target at a rate of 64.9%, the intermediate group at 62.5%, and the advanced group at 68.6%. Due to the close proximity of these target levels, this factor group was not considered a significant predictor of accuracy.

Despite the fact that the level of 'error'[9] is relatively similar among the groups, a closer look at the data reveal that there are in fact clear differences in error type between each group; while the number of errors of failing to voice in voicing contexts increases with speaker experience, the number of errors of voicing in non-voicing contexts decreases with experience.

Table 8. Non-target sibilants produced by speaker level



The novice participants committed a total of 60 errors, with 37 (61%) of them resulting from failing to produce regressive voicing assimilation (labeled 'non-target [s]' in Table 8) in the context that calls for it and 23 (38%) from voicing in a non-voicing context (labeled 'non-target [z]'). The intermediate group made a total of 63 errors, 44 (69%) of them being non-target [s] and 19 (30%) being non-target [z]. The advanced group committed a total of 55 errors, with 51 (92%) of them being non-target [s] realization while only 4 of them (7%) were non-target [z] errors. These results suggest that as student exposure, experience, and proficiency increases, they may become more aware of the restriction that Spanish has on intervocalic voicing and tend to avoid it. However, they still fall short in terms of producing regressive voicing assimilation and instead seem to be hypercorrecting to avoid [z] in any context.

Status of cognate. Overall, there were 270 target words that were cognates and 195 that were non-cognates. Within the group of cognates, 127 of those contained a voiced sibilant in the English equivalent (e.g. *visita* 'visit') and 143 contained a voiceless sibilant in the English equivalent (e.g. *clase* 'class'). The data reveal that participants produced nearly categorical voicelessness when the English equivalent was voiceless, with only 1 token out of the 143 being voiced. However, when the English equivalent is voiced, the voicing rate jumps to 26.2%—a number much higher than the overall voicing rate of 12.5%. Of the tokens that are not part of a cognate, the rate of voicing (9.2%) is closer to the overall voicing rate. This suggests that phonological transfer among cognates may be stronger than that of non-cognates. The Rbrul weights confirm a strong correlation between cognate status and voicing; cognates with [s] in English nearly categorically condition [s] in Spanish with a weight of 0.913 while cognates with [z] in English highly inhibit [s] in Spanish with a weight of 0.136. Non-cognates, with a weight of 0.377, also condition [z].

As seen in Table 5, among cognates that are voiced in English, there is nearly a 50/50 split between target and non-target production with non-targets marginally outpacing targets 51.2% to 48.8%. The rate of hitting the target among cognates with [z] in English is lower than the overall rate of target production of 65.4%. A closer look at the word list and the overall patterns found in the data relating to following phonetic context gives insight into why

this may have occurred: within the group of cognates with voicing in English, there were words where /s/ was in intervocalic position (*visita* 'visit'*, música* 'music'*, president* 'president') as well as when /s/ was followed by a voiced consonant (*organismo* 'organism'*, turismo* 'tourism'*, mecanismo* 'mechanism'). We know that participants often produced errors with /s/ in intervocalic position due to voicing and produced errors in the /VsC$_{voiced}$/ contexts due to lack of voicing, and therefore the [z] cognate factor witnessed errors of both types. Participants hit the target sound at a rate of 84% among cognates that are voiceless in English such as *disco* 'disc', *respeto* 'respect', and *pasar* 'to pass'. This is a higher rate of accuracy than the overall rate of 65.4%; an intuitive result considering factors of phonological transfer; if the participant were to produce [z], he/she would have to produce the opposite sound from what is used in equivalent word in the L1. The Rbrul weights in Table 5 confirm the tendencies of cognates with voicing in English to condition the production of the non-target sound with a weight of 0.717 and of cognates with voicelessness in English to inhibit the non-target with a weight of 0.325. Non-cognates fell in middle ground, with a slight favoring of the non-target at 0.451.

Knowledge of word. Overall, participants were familiar with 92.8% of the words given to them. Of the words that they did know, they produced voicing at a rate of 13.3%, a rate very similar to the overall rate of voicing of 12.5%. Of the words that they did not know, however, they produced very low rates of voicing—only 1 token out of 37 (2.7%). This suggests that when a word is unknown, speakers may be even less likely to produce voicing than when the word is known. The Rbrul weights illustrate this distinction with known words favoring [z] to a high degree at 0.743 and unknown words inhibiting [z] at 0.257[10].

Considering target/non-target as the dependent variable, participants were twice as accurate within the known words at 68.1% as compared to unknown words (29.7%). However, the factor group was ruled as an insignificant predictor for this variable.

5. Discussion

Overall, the participants differed substantially in production of /s/ as compared to the monolingual speakers reported in Schmidt & Willis (2011). There were two main differences: participants produced [z] instead of [s] in intervocalic position at an overall rate of 18% and produced [s] instead of [z] when followed by a voiced consonant at a rate of 87%. As experience and proficiency increased, intervocalic voicing decreased but the failure to regressively assimilate /s/ increased.

One possible interpretation of this finding can be seen as a validation of Major's (2014) Ontogeny and Phylogeny Model of second language acquisition. According to this model, there are two types of error in an interlanguage grammar: transfer errors and developmental errors. Transfer errors are those that come as a result of the rules of the L1, while developmental errors involve the same sort of errors that children make in acquiring their L1 such as overgeneralizing a specific rule. The model predicts that the kind of errors made by second language learners will be dependent on their level of proficiency. Novice students may have nothing to draw on but their L1, and therefore will make a significant amount of transfer errors that will decrease over time. However, as speakers progress in proficiency, they acquire a certain amount of knowledge about the L2 and this knowledge becomes a potential source of developmental error. Developmental errors begin low among novice speakers, increase among intermediates, and then decrease again among advanced.

Applying this model to the present data, errors in intervocalic voicing would be considered a transfer error, which would decrease with speaker proficiency. Failing to produce regressive voicing assimilation in the appropriate context can be considered a developmental error that is uncommon at the novice level but that increases with proficiency. Even

though the model predicts that developmental errors will return to a low level with high proficiency speakers, it also acquiesces that not all learners reach the advanced stage in all areas. For example, Archibald (1998, p. 5) states that it is common for learners to reach a plateau in their acquisition. He gives the example that even after many years of exposure to English, a second language learner may still produce sentences like "I don't know what should I do.". Additionally, just because a particular learner commits developmental errors does not mean that he/she has not reached the advanced speaker range, but rather has not acquired advanced accuracy in that one particular area (in this case, acquisition of /s/).

A second interpretation of this data is that a lack of assimilation may not be perceived to be an error at all. Because native-speakers produce regressive voicing assimilation variably as well as gradiently, they may accept both variants— [s] and [z] in /VsC$_{voiced}$/ contexts (as well as combinations of both) —as equally valid. In that case, the advanced speakers of this study would have low levels of non-target realizations, suggesting that they have fully acquired the phonological rules governing Spanish /s/. Future studies that examine the perception of Spanish /s/, specifically addressing if a lack of voicing contributes to a foreign accent, would clarify the results of the present investigation.

The second research question tackles the issue of what linguistic and extralinguistic factors constrain the production of non-target sibilants. The fact that following phonetic context was an insignificant predictor of voicing suggests that L2 speakers do not follow the native speaker patterns in terms of regressive voicing assimilation.

Cognates were found to impact /s/ production among our participants, with cognates that contained a voiced sibilant in English highly inhibiting [s] realization and those that contained a voiceless sibilant highly conditioning [s]. Unfortunately, it was outside of the scope of the study to investigate how different types of cognates, such as those that are similar orthographically versus those that are similar phonetically, might affect production, as well as how the grade of difference between the cognates could also play a role.

Orthography was not found to be a significant predictor of voicing/voicelessness, nor of target/non-target sibilant production. This is in contrast to what Miglio and Fukazawa (2006) propose, "that spelling interference is a powerful force in the shaping of the early interlanguage phonology and should be taken into account in the formal analysis of these phenomena" (p. 4143). Further studies that investigate the role of orthography, especially in the case of Spanish /s/, are needed.

6. Conclusion

This small-scale, preliminary study was designed to shed light on the acquisitional process of Spanish /s/ among L1 English speakers, with the goal of understanding to what extent L2 speakers comply with the prescriptive norms of regressive voicing assimilation and intervocalic voicelessness. The data suggest that participants, regardless of proficiency level, prefer to maintain the voiceless variant [s] preceding a voiced consonant rather than assimilating to that consonant in the [voicing] feature. At the same time, novice and intermediate speakers also tend to produce intervocalic voicing, a characteristic typically absent from native-speaker speech[11], although by the advanced level it almost disappears entirely. Given the variable and gradient nature of assimilation, it is unclear if failing to assimilate would be perceived as non-native and therefore even considered a non-target sound.

Due to the limited scope of the study, not all factors associated with the acquisition of Spanish /s/ were taken into consideration. As Schmidt (2014) noted in her study, future work is required to determine what other linguistic variables (e.g. speech rate, continuity of speech, manner of articulation of following consonant) and learner factors (e.g. type and degree of experience with Spanish, role of phonetics

training) may play a role. Lastly, future studies that examine the perception of Spanish /s/ as it relates to foreign accent would be valuable for understanding how the non-target sibilants produced by the participants of the current study are perceived. Textbook descriptions can predict if a sibilant is target or not, however, the judgment of both native speakers as well as fellow L2 speakers provides invaluable information regarding what learners hear as foreign-accented and what they perceive and do not perceive in the speech signal (Schoonmaker-Gates, 2012).

Notes

1. IRB approval date: 1/29/15
2. Schmidt & Willis (2011) found notable differences between historical descriptions and actual phonetic realization regarding /s/ when followed by voiced consonants. Their study revealed that in a monolingual sample of Mexico City speakers, 37% of tokens were produced without voicing in the contexts that are expected as voicing triggers.
3. Not taking into account other possible pronunciation errors.
4. Mexico City was chosen because of its status as a non-/s/-aspirating dialect
5. Table 2 in Schmidt and Willis (2011) has the first two phonetic contexts reversed but the following segments and examples listed the same as this table; I assume this to be a small typo so I have corrected them here.
6. At the 0.05 level.
7. Pseudonyms have been given to protect anonymity.
8. This factor group was included as a random effect during the Rbrul run for reasons of statistical validity.
9. 'Error' is defined here as a descriptively non-target sound, however, the researcher urges caution with such an interpretation, given that it is unclear that native-speakers would classify lack of voicing as a characteristic of non-native speech.
10. However, the fact that there was only 1 token of voicing among the unknown words, along with the fact that there were only 37 unknown words out of 511 makes it difficult to draw reliable conclusions from this factor group.
11. For an exception, see Lipski's (1989) account of Andean intervocalic voicing.

Works Cited

Al-Harbi, A. (2005). English Voicing Assimilation: Input -to-Output [±voice] and Output-to-Input [±voice]. *JKAU Arts & Humanities*, *13*, 61-81.

Archibald, J. (1998). *Second language phonology*. Amsterdam: Benjamins.

Bayley, R. & Holland, C. (2014). Variation in Chicano English: The case of final (z) devoicing. *American Speech*, *89*(4), 385-407.

Browman, C. & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219-252.

Cobb, K. & Simonet, M. (2015). Adult Second Language Learning of Spanish Vowels. *Hispania*, *98*(1), 47-60.

Cotton, E. G. & Sharp, J. M. (1988). *Spanish in the Americas*. Washington, D.C.: Georgetown U Press.

Díaz, M. (2011). Acento anglo en español: fuentes y posibles soluciones. Quebec Aprende Español. University of Montreal.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, *26*, 249–273.

Dykstra, G. (1955). *Spectrographic analysis of Spanish sibilants and its relation to Navarro's physiological phonetic descriptions*. Unpublished doctoral dissertation. University of Michigan.

Face, T. L. (2006). Intervocalic rhotic pronunciation in adult learners of Spanish as a second language. In C. Klee & T. Face (Eds.), *Selected Proceedings of the Seventh Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, (pp. 47–58). Somerville, MA: Cascadilla.

Face, T. L. & Menke, M. (2009). Acquisition of the Spanish voiced spirants by second language learners. In J. Collentine (Ed.), *Selected Proceedings of the Eleventh Hispanic Linguistics Symposium*, 39–52. Somerville, MA: Cascadilla.

García, A. (2013). *Allophonic Variation in the Spanish Sibilant Fricative*. Doctoral Dissertation. The University of Wisconsin, Madison.

González-Bueno, M. (1995). Adquisición de los alófonos fricativos de las oclusivas sonoras españolas por

aprendices de español como segunda lengua. *Estudios de Lingüística Aplicada, 13*, 64–79.

Gorman, K. & Johnson, D. E. (2013). Quantitative analysis. In R. Bayley, R. Cameron, & C. Lucas (Eds.), *The Oxford Handbook of Sociolinguistics*, (pp. 214–240). Oxford: Oxford University Press.

Guitart, J. (2004). *Sonido y sentido:* teoría y práctica de la pronunciación del español contemporáneo con audio CD. Washington, DC: Georgetown University Press.

Hualde, J. I. (2005). *The Sounds of Spanish*. Cambridge: Cambridge University Press.

Jiménez-Sabater, M. (1975). *Más datos sobre el español de la Republica Dominicana*. Santo Domingo: Ediciones Intec.

Johnson, D. E. (2009). Getting off the Goldvarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 3, 359–383.

Knight, R. A. (2003). Understanding English Variation: Week 4 – Connected speech processes and the principles of English spelling. Lecture notes. University of Surrey, Roehampton. www.rachaelanne.net/teaching/uev/uev4.doc.

Lipski, J. (1989). /s/-Voicing in Ecuadorian Spanish: Patterns and principles of consonantal modification. *Lingua*, *79*, 49-71.

Magloire, J. & Green, K. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica*, *56*, 158–185.

Major, C. R. (1986). The ontogeny model: evidence from L2 acquisition of Spanish r. *Language Learning, 36*, 453–504.

Major, C. R. (2014). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. New York: Routledge.

Menke, M. & Face, T. (2010). Second language Spanish vowels production: An acoustic analysis. *Studies in Hispanic and Lusophone Linguistics*, *3*, 181-214.

Miglio, V. & Fukazawa, H. (2006). Spelling Influence on Pronunciation: Output-Output Correspondence and Interlanguage Phonology (pp. 4143-4152). Proceedings from the 4th Hawaii International Conference on the Arts and Humanities.

Obaid, A. (1973). The vagaries of the Spanish 's'. *Hispania*, *56*(1), 60-67.

Quilis, A. (1993). *Tratado de fonología y fonética españolas*. Madrid: Editorial Gredos.

Reeder, J. (1998). English speakers' acquisition of voiceless stops and trills in L2 Spanish. *Texas Papers in Foreign Language Education, 3*, 101–118.

Robinson, K. (1979). On the voicing of intervocalic s in the Ecuadorian highlands. *Romance Philology*, *33*(1), 137-143.

Romero, J. (1999). The effect of voicing assimilation on gestural coordination. *Proceedings of the XIVth International Congress of Phonetic Sciences*, (pp. 1793-1796). San Francisco, CA.

Rose, M. (2010). Differences in discriminating L2 consonants: a comparison of Spanish taps and trills. In M. Prior, Y. Watanabe, & S. Lee (Eds.), *Selected Proceedings of the 2008 Second Language Research Forum*, (pp. 181–196). Somerville, MA: Cascadilla.

Sankoff D., Tagliamonte S.A., & Smith E. (2012). GoldVarb LION: A variable rule application for Macintosh, Department of Linguistics, University of Toronto, and Department of Mathematics, University of Ottowa.

Schmidt, L. (2011). *Acquisition of dialectal variation in a second language: L2 perception of aspiration of Spanish /s/.* Indiana University, Bloomington, Indiana. Doctoral dissertation.

Schmidt, L. (2014). Contextual Variation in L2 Spanish: Voicing Assimilation in Advanced Learner Speech. *Studies in Hispanic and Lusophone Linguistics*, *7*(1), 79-114.

Schmidt, L. & Willis, E. (2011). Systematic Investigation of Voicing Assimilation of Spanish /s/ in Mexico City. In S. M. Alvord (Ed.), *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology*, (pp. 1-20). Somerville, MA: Cascadilla Proceedings Project.

Schoonmaker-Gates, E. (2012). *Perception of foreign accent in Spanish by native and nonnative listeners: Investigating the role of VOT and speech rate*. Doctoral dissertation. Indiana University, Bloomington, Indiana.

Schwegler, A., Kempff, J. & Ameal-Guerra, A. (2010). *Fonética y fonología españolas: Teoría y práctica (4th ed.).* New York: John Wiley & Sons, Inc.

Torreblanca, M. (1978). El fonema /s/ en la lengua española. *Hispania*, *61*(3), 498-503.

Torreblanca, M. (1986). Prevocalic voiced s in modern Spanish. *Thesaurus*, *41*(1-3), 59-69.

Waltmunson, J. (2005). *The relative degree of difficulty of Spanish/t, d/, trill and tap by L1 English speakers: Auditory and acoustic methods of defining pronunciation accuracy*. Doctoral dissertation. University of Washington.

Whitley, M. S. (2002). *A course in Spanish linguistics: Spanish/English contrasts (2nd ed.)*. Washington, DC: Georgetown University Press.

Zampini, M. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania,* 77, 470–481.

Zampini, M. (1998). The relationship between the production and perception of L2 Spanish stops. *Texas Papers in Foreign Language Education*, *3*, 85–100.

Zampini, M. & Green, K. (2001). The voicing contrast in English and Spanish: the relationship between perception and production. In J. Nicol (Ed.), *One mind, two languages: bilingual language processing*, (pp. 23–48). London: Blackwell.