1. Introduction

One of the generally accepted phonological processes in Spanish phonology is that of assimilatory voicing of /s/ before a voiced consonant. This process is taught in a near categorical fashion in most phonetics text books written for second language learners in the United States (e.g., Quilis and Fernández 1985). However, to our knowledge there has been no systematic examination of this voicing process in Spanish, rather a few anecdotal and impressionist claims. This lack of research on voicing of /s/ in coda position is also somewhat remarkable given the literature on /s/ in aspirating dialects. We chose to first examine one of the putative non-aspirating dialects of Spanish to ensure that our characterization was not misinterpreted by debuccalization or aspiration. Our principal goal is to provide an initial examination in a systematic fashion that could be used both as a dialectal characterization and to understand the process. We specifically used a controlled task of repeated tokens in a variety of contexts across a relatively homogeneous informant group that was recorded in the target dialect, Mexico City. The paper proceeds (Section 2) with a review of the literature on voicing assimilation and the few anecdotal claims of variation. Section 3 provides the methodology of the current study, including informants, instrument, and equipment. The findings are presented in Section 4 and then discussed in Section 5. We provide our conclusions in Section 6.

2. Normative Spanish /s/ and Voice Assimilation

2.1 Voice Assimilation of Spanish /s/

At one point in the history of Spanish there was a phonemic voiced/voiceless alveolar fricative distinction that was lost during the early Modern Spanish beginning in the 16th century. This phonemic contrast was likely neutralized in coda position. Penny asserts (1993: 80), “it is highly likely that the medieval sibilants adopted the voice-quality of the following syllable initial consonant (where this existed, either in the same word as the sibilant or at the beginning of the following word).” The voicing contrast does not contribute to a phonemic distinction in Modern Spanish, though it did remain in other Romance Languages such as Portuguese, Catalan, French, among others (Harris and Vincent 1988).

Allophonic variation of the voiceless /s/ can result in a voiced variant in particular contexts. In their classic textbook for English-speaking learners of Spanish and referring to the linguoalveolar fricative /s/ in contact with a voiced consonant, Quilis and Fernández (1985) state that /s/ is almost always produced as a voiced variant. The authors maintain that this voicing is due to lack of control of the vocal folds, with transmission of voicing of the neighboring voiced consonant to the voiceless consonant, /s/. However, they also state that this production is not consistent1.

* We would like to acknowledge and express our gratitude to Elizabeth Santana Cepero of the Universidad Autónoma Metropolitana-Iztapalapa for providing access to recording facilities and her students.

1 Referring to the the allophone [z], “Se produce casi siempre que el fonema /s/ precede a una consonante sonora, ya que, entonces, por descontrol de las cuerdas vocales, se transmite la sonorización a la consonante sorda. Esta realización no es constante” (1985: 97).
Allophonic complementary distribution of the voiced and voiceless sibilants is specified in contexts preceding a voiced consonant. This regressive assimilation in voice of /s/ to the following voiced consonant is described in a number of introductory texts, although it is often presented as a tendency rather than a mandatory process (e.g., Hammond 2001; Hualde 2005; Schwegler et al. 2010; Whitley 2002). This process is claimed to occur both word-internally as well as across word boundaries and is exemplified in (1). Preceding a voiceless consonant, /s/ is predicted to remain voiceless, as in después. However, /s/ preceding a voiced consonant, such as in desde and los bancos, may be realized with voicing.

(1) después → [des.pwes] ‘after’ (/s/ is not voiced)
desde → [dez.de] ‘since’ (/s/ is voiced word-internally)
los bancos → [loz.βaŋ.kos] ‘the banks’ (/s/ is voiced across word boundary)

Quilis (1993) states that /s/, like any voiceless consonant, may be produced as voiced, most of all, when it is found in contact with a voiced consonant (251). When there is voicing of the /s/ segment, Quilis understands this to be due to a lenition process (1993: 67, 251). As the context that predicts /s/ voicing is triggered by the voicing quality of the following consonant, (2a) provides a list of those contexts where voicing of /s/ can occur, while (2b) lists those contexts in which voicing is not predicted (Hualde 2005: 160):

(2a) voiced stops: s.b s.d s.g
nasals: s.m s.n
liquids: s.l s.r
(glides: s.y s.w
fricatives: s.f s.s s.x
affricates: s.tʃ
vowels\(^3\): s.V.

An example of voicing assimilation from the current corpus is shown below in Figure 1.

---

**Figure 1.** Example of phonemic /s/ realized as [z] in the name Esmeralda.

---

2 The /s.r/ pair often undergoes a double assimilation in which the /s/ becomes voiced and the /r/ is assimilated, resulting in both a regressive assimilation and progressive assimilation (Quilis 1993: 81, 251).

3 Voicing of /s/ does not typically occur between vowels in Spanish, except in some limited dialects such as the highlands of Ecuador (Lipski 1989), where word-final prevocalic /s/ is voiced.
There have been a number of reports on coda /s/ voice assimilation, and they suggest that the process may be variable (Torreblanca 1978, 1986; Obaid 1973; Dykstra 1955). Voice assimilation has been reported in central Peninsular varieties (Torreblanca 1978, 1986); Monterrey, Mexico (Obaid 1973); and Highland Ecuador (Lipski 1989, Robinson 1979). Dykstra (1955) also reports voicing throughout Latin American dialects, including countries where aspiration is present.

These accounts, however, 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. The variable descriptions of voicing assimilation are presented in Table 1.

<table>
<thead>
<tr>
<th>Phonetic context of voicing of /s/</th>
<th>Voicing before a voiced consonant only (Schwegler et al 2010)</th>
<th>Cases of voicing before a voiceless consonant, vowel, or pause (Robinson 1979, Torreblanca 1978, Obaid 1973)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direction of assimilation</td>
<td>Regressive assimilation (Schwegler et al 2010)</td>
<td>Progressive assimilation (Dykstra 1955)</td>
</tr>
<tr>
<td>Motivation</td>
<td>Due to syllable-final articulatory relaxation (Torreblanca 1978)</td>
<td>Due to articulatory gesture overlap (Quilis 1993, Hualde 2005: 115)</td>
</tr>
<tr>
<td>Regional variation</td>
<td>Occurs only in /s/-conserving varieties (Cotton &amp; Sharp 1988)</td>
<td>Cases of voiced aspirated-/s/ in /s/-weakening variety (Jimenez-Sabater 1975, Dykstra 1955)</td>
</tr>
</tbody>
</table>

Table 1. Descriptions and proposals of Spanish /s/ voicing.

2.2. Dialect Characterization of Mexican /s/

The production of the segment /s/ in Mexican Spanish is widely recognized for its full retention of /s/ in non-coastal dialects. Henríquez-Ureña noted in 1921 that, “the frication of the Mexican /s/ is very long, and moreover they are pronounced more or less equally at the beginning and end of the syllable.” He later reported “The consonants are of a long tension and of a precise articulation: it is well known that the /s/ of Mexico, [is characterized as] dental, upon the lower incisors, of a sharp timber, singular for its duration among all [the dialects] of the Hispanic world” (translation our own).

Concerning the production of /s/ in the central valley, or greater Mexico City region, Moreno de Alba (1994) notes (translation our own):

> Of what concerns the mode and point of articulation, the studies are in agreement in that the allophone that is most frequent, in all the country, is the predorsal voiceless fricative, as Perissinotto (1975) defines for the city of Mexico: “its articulation is predorsal fricative voiceless [s], with the tongue in a convex position and in contact with the posterior part of the lower incisors” (56).

Voicing of intervocalic /s/ has been reported within Mexican Spanish. From a sociolinguistic perspective, Martín-Butragueño (in press) reviews the allophonic variants of /s/ in intervocalic position in the Spanish of Mexico based on the data in the Atlas Lingüístico de México. One of the features

---

4 “La fricación de la s mexicana es muy larga, y además se pronuncia más o menos igual en principio y a fin de sílaba” (Henríquez Ureña 1921:25), (cited in Moreno de Alba 1994:101).

5 “Las consonantes son de larga tensión y de articulación muy precisa: es bien conocida la s mejicana, dental, apoyada en los incisivos inferiores, de timbre agudo, singular por su longitud entre todas las del mundo hispánico” (Henríquez Ureña 1938: 336).

6 “Por lo que toca al modo y punto de articulación, los estudios están de acuerdo en que el alófono más frecuente, en todo el país, es el predorsal fricativo sordo, como bien define Perissinotto (1975) para la ciudad de México: ‘su articulación es predorsal fricativa sorda [s], con la lengua en posición convexa y en contacto con la parte posterior de los incisivos inferiores’ (56)” (100-101).
considered was the presence of voicing. He reports that of 1,329 annotations of total productions representing 29 different localities, there were 159 reports of voiced tokens in the collection. Voiced tokens of intervocalic /s/ were not limited to specific geographic regions, but were found throughout the country (p. 23), albeit at low rates.

2.3. Research Questions

The review of literature with conflicting claims of voicing assimilation as well as suggestions of variability across dialects, compounded with the lack of empirical data on the topic, leads us to the following research questions:

1. What is the nature of the process of /s/ assimilatory voicing in Spanish? Are systematic patterns observed within a specific dialect of Spanish?
2. How do the linguistic factors following segment and phrase position influence the production of Spanish /s/ voicing?
3. Is variation found in voicing of Spanish /s/ according to sex? What is the role of the individual speaker?

3. Methodology

3.1. Informants

The informants for the study were twelve university students attending the Universidad Autónoma Metropolitana-Iztapalapa (UAM-I), 6 females, 6 males. The dialect region was the greater Mexico City metropolitan area. The informant ages ranged from 21-29, with a mean age of 25. The majority of the speakers had lived their entire lives in the state of Mexico (central region). Three of the 12 speakers came to Mexico City between the ages of 9-14, but they came from neighboring states. As the informants were attending university they had some knowledge of foreign languages (English or French), but did not consider themselves fluent. The informants were recruited by social networks or by invitation from the second author while on the UAM-I university campus.

3.2. Data Elicitation

3.2.1. Recording

The data was recorded during June of 2010 by the second author in the sound attenuated phonetics laboratory at the Universidad Autónoma Metropolitana-Iztapalapa. The data was recorded with a Shure WH20 head-mounted microphone connected to an USBPre external sound card and then fed into a laptop computer. The sound was captured with CoolEdit 2000.

The recording session began with a preliminary sociolinguistic interview to familiarize the informants with the microphone and to verify the dialectal origin of the speaker. The informants then completed a series of three tasks. The first task completed was a contextualized picture-description task presented in a PowerPoint presentation, followed by a 200-word story reading task and comprehension quiz. The final task completed by the informants was a written language background questionnaire, which elicited information regarding the language(s) spoken by the informant, demographic information (age, sex, profession), places lived and traveled, and parents’ native language(s) and origin. The current analysis considers only the data elicited from the contextualized picture-description task and the language background questionnaire.

---

7 In a personal communication, Martín-Butragueño explains that the 1,329 annotations in the Atlas Lingüístico de México are a sum of each of the different variants found in each locality (see Martín-Butragueño in press).
8 We have included this variable, which has not been previously addressed in research on /s/-voicing, because studies of a sociolinguistic nature have often found differences in sex in the realization of a particular variant (e.g., Serrano 2006, vowel reduction in Mexican Spanish).
9 This is based on a self-report in their background questionnaire.
3.2.2. Instrument

The picture-description task was designed in order to elicit semi-directed continuous speech in the form of phrase-long utterances. Through the use of this task, the same contexts were elicited across informants in a more natural way than a simple reading task. The picture-description task, administered via PowerPoint, presented scenes on individual slides of a street market vendor in the Mercado Santiago with one of their wares for sale.

Before beginning the descriptions in the Mercado, informants first completed a training session in which they were familiarized with the pictures and vocabulary items used in the second part of the task and were trained on how to complete the task. Two or three letters of the beginning of the targeted words were presented along with the pictures of the objects in the vocabulary naming practice so as to encourage the informants to produce the desired lexical items. This training allowed us to verify the production of particular lexical items.10

The second half of the task presented individual slides that contained vendor names and objects that the vendor sold. For each slide, the informants were asked to state the vendor’s name, what he or she sold, and the number of objects to an imaginary person on the phone (for example, Luis vende siete revistas). The names of the vendors and the objects were chosen specifically to present both word-internal and word-final contexts of /s/ followed by voiced and voiceless consonants and by a vowel. The vendors’ names (but not the names of the objects being sold) were presented visually on the slides. Each name was presented two times throughout the task while the objects were seen only once. The informants were given five seconds per slide to report on the vendor’s stall before automatically advancing to encourage a consistent speech rate. There were a total of 56 slides, and thus 56 elicitations of the phrase “Name vende (*sells*) number object”. The informants were offered a short optional break approximately halfway through the task.

The current analysis is limited to the context of word-internal /s/. 11 There were a total of 49 possible tokens of word-internal /s/ per informant, provided the informant produced all of the elicited names and objects in the production task. Three types of word-internal phonetic contexts were included: (1) /s/ preceding a voiced consonant, (2) /s/ preceding a voiceless consonant, and (3) /s/ preceding a vowel (intervocalic /s/). Although an attempt was made to include all consonant types for following context, contexts were limited to voiceless stops, nasals, the voiced bilabial stop, the lateral liquid, and the mid and low vowels. This is due to the low frequency of word-internal /s/ preceding other consonant types in Spanish and the restriction of the task to lexical items that could be represented by images and sold in a market. Table 2 presents the distribution of token types according to the phonetic context. See the appendix for the complete list of contexts elicited.

<table>
<thead>
<tr>
<th>Phonetic context</th>
<th>Following segment</th>
<th>Examples</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>/VsC&lt;sub&gt;voiced&lt;/sub&gt;/</td>
<td>/p, t, k/</td>
<td>Oscar, espejos, postres</td>
<td>17</td>
</tr>
<tr>
<td>/VsC&lt;sub&gt;voiceless&lt;/sub&gt;/</td>
<td>/b, l, m, n/</td>
<td>Osvaldo, isla, cisne</td>
<td>15</td>
</tr>
<tr>
<td>/VsV/</td>
<td>/a, e, o/</td>
<td>César, narices, payasos</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>49</td>
</tr>
</tbody>
</table>

Table 2. Contexts elicited in picture-description task for each speaker.

3.3. Analysis

For each of the twelve informants, all tokens of word-internal /s/ were extracted from the larger speech sample elicited in the picture-description task, for a total of 567 tokens. Using Praat software (Boersma and Weenik 2010, version 5.1.30) for acoustic analysis, measurements of the duration of voiced sibilance and of voiceless sibilance were made for each token, as well as the duration of the

10 Despite the lexical training a few words were produced with a local vocabulary item such as postre instead of pastel ‘cake’, or tapete instead of alfombra ‘carpet’. These variations were few and were incorporated into the data where possible.

11 For the current study we only examine the word-internal tokens. Tokens from the instrument with voicing across word boundaries will be analyzed in future research.
total /s/ segment. The sibilant segment was defined by the visual presence in the spectrogram of strong high-frequency frication, around 8000 to 9000 Hertz. Voicing was determined by the presence of (1) visible regular glottal pulses in the voice bar and (2) regular periodic patterns in the waveform.

![Figure 2](image)

*Figure 2*. Waveform, spectrogram and textgrid of example file for coding. Numbers above the waveform indicate the boundaries of specific elements of the segment /s/.

A percentage voiced was calculated for each token, by dividing the duration of voiced sibilance by the total duration of the sibilant segment (duration between arrows 1 and 2 divided by the duration of arrows 1 and 3 illustrated in Figure 2). Those tokens of /s/ in which the majority of the segment was voiced (60-100%) were coded as voiced, while tokens in which the majority of the segment was voiceless, or was made up of less than 40 ms of voicing, were coded as voiceless.\(^\text{12}\) Coding was performed by both authors and validated in meetings as well as through an inter-rater analysis of 20% of the data.\(^\text{13}\) Each token was coded according to: (a) voicing of /s/ (voiced or voiceless), (b) position of /s/ within the phrase (phrase-internal or phrase-final), (c) following segment type (vowel, voiced consonant, or voiceless consonant), (d) sex of the speaker, and (e) individual speaker. The decision to impose a categorical judgment, voiced versus voiceless, on a phenomenon that presents gradient realizations was made to facilitate comparisons with previous literature and is based on general tendencies of voicing observed in the data. The categories of voiced or voiceless also permit us to see patterns due to the independent variables.\(^\text{14}\)

\(^{12}\) The range of progressive voicing from the vowel onto the sibilant was 0-34 ms in intervocalic position an up to 37 ms preceding a voiceless consonant. We chose 40 ms as an arbitrary cut-off point due to these patterns in our data in the expected voiceless context, [VsV]. Because some tokens were very short, a strict percentage judgment would inflate the distribution in favor of voiced when the actual voicing present is normal for voiceless tokens. This metric was used in only a few cases.

\(^{13}\) All cases that were problematic were analyzed by both authors until a consensus was achieved.

\(^{14}\) While this categorization is somewhat arbitrarily set at 60-100% voiced, our documentation of the boundaries of the categories allows for replication and comparison.
The voicing percentage as presented reflects left-edge voicing, a continuation of the voicing of the preceding vowel onto the following sibilant. We did not calculate right-edge voicing due to its very infrequent presentation in the data. We started the analysis with the assumption that there would be both left-edge and right-edge voicing. However, since right-edge voicing separate from total voicing of the segment was so infrequent in our corpus, it was eliminated from the calculation.

Statistical analyses using SPSS software were undertaken to determine significant main effects and interactions of these variables on voicing of the sibilant. Linear mixed models with Bonferroni adjustments for multiple comparisons were conducted and are further discussed in Section 4.5, following the presentation of the findings observed from the initial descriptive analyses. The dependent variable tested in the statistical analyses was percentage voiced (0-100%), rather than the categorical voiced versus voiceless, to facilitate statistical analyses and to reflect the gradient nature of the voicing in the data.

4. Findings

4.1. Distribution of Voiced [z] and Voiceless [s] Variants

The distribution of the voiced and voiceless realizations of /s/ is presented, according to the contexts where voicing assimilation is expected. Table 3 presents a frequency count of the voiced and voiceless tokens for the group of Mexico City speakers for each of the three phonetic contexts: /s/ preceding a voiceless consonant, preceding a vowel, and preceding a voiced consonant.\textsuperscript{15}

<table>
<thead>
<tr>
<th></th>
<th>Non-voicing context</th>
<th>Voicing context</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/VsC\textit{voiceless}/ (N=187)</td>
<td>/VsV/ (N=214)</td>
</tr>
<tr>
<td>Voiced [z]</td>
<td>5% (10)</td>
<td>9% (20)</td>
</tr>
<tr>
<td>Voiceless [s]</td>
<td>95% (177)</td>
<td>91% (194)</td>
</tr>
</tbody>
</table>

Table 3. Frequency count of voiced and voiceless variants according to phonetic context, (N tokens).

The most striking fact in Table 3 is the percentage of tokens, 37%, without voicing in the contexts that are expected as voicing triggers. While we see a trend for voicing of /s/ preceding a voiced consonant, it is also clear that the voicing assimilation process is far from categorical in its occurrence. Also in Table 3, we see that not all realizations of /s/ preceding a voiceless consonant or preceding a vowel were voiceless. 5% of the tokens of /s/ before a voiceless consonant and 9% of the tokens of /s/ in intervocalic position were voiced.

4.2. Phrase Position

Next, the effect of position within the phrase on voicing of /s/ is considered. The contexts of word-internal /s/ were divided into two phrase positions depending upon where they occur in the utterances produced: phrase internal and phrase final\textsuperscript{16}. Figure 3 presents the proportion of tokens that are realized as voiceless [s] and as voiced [z] according to the following segment for each phrase position.

\textsuperscript{15} Statistical analyses of these results are presented in Section 4.5.

\textsuperscript{16} We use the term phrase final because the phonological process occurs in the final word of the utterance. The use of this term does not imply that the assimilation is the final sound of the utterance.
As observed in Figure 3, voicing of /s/ in the expected voicing context, preceding a voiced consonant, is realized more frequently when it occurs in phrase-internal position (in 76.4% of the contexts) than in phrase-final position (in 53.2% of the contexts). Only two of the twelve speakers did not follow this pattern (speakers 6F and 7M). An example of a voiceless production of /s/ preceding a voiced consonant in phrase-final position is displayed in the token *chisme* (‘gossip’), Figure 4.
4.3. Sex

4.3.1. Voicing According to Sex

After examining the effect of the linguistic variables following segment type and phrase position, we now turn to consider the extra-linguistic variables sex and individual speaker. Table 4 presents the distribution of voiced and voiceless realizations of /s/ in each of the three phonetic contexts for the female (N=6) and male (N=6) groups. As observed in Table 4, the male speakers appear to voice /s/ in the predicted voicing context to a slightly greater degree than the female speakers (67% versus 59%). Also, the majority of the cases of voicing of /s/ preceding a voiceless consonant were realized by the female speakers (speakers 2F, 9F, and 10F), while most cases of voicing of intervocalic /s/ were produced by the male speakers (speakers 12M and 3M).

<table>
<thead>
<tr>
<th></th>
<th>/VsC_{voiceless}/ (N=187)</th>
<th>/VsV/ (N=214)</th>
<th>/VsC_{voiced}/ (N=166)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Voiceless</strong></td>
<td>82 (90%)</td>
<td>104 (97%)</td>
<td>34 (41%)</td>
</tr>
<tr>
<td><strong>Voiced</strong></td>
<td>9 (10%)</td>
<td>3 (3%)</td>
<td>49 (59%)</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Males</strong></td>
<td>95 (99%)</td>
<td>90 (84%)</td>
<td>27 (33%)</td>
</tr>
<tr>
<td></td>
<td>1 (1%)</td>
<td>17 (16%)</td>
<td>56 (67%)</td>
</tr>
</tbody>
</table>

*Table 4. Distribution (frequency count) of voiced and voiceless variants according to following segment and sex.*

4.3.2. Voicing According to Sex and Phrase Position

An interesting difference in voicing according to sex is observed when the phrase position is considered (Figure 5). While both the female and male speaker groups average voicing in 76%-77% of the predicted contexts phrase-externally, a difference in frequency of voicing according to sex appears in phrase-final position. The female speakers voice the sibilant less often than the male speakers at the end of a phrase (with 46% compared to 61% voicing in the relevant contexts, respectively).

*Figure 5. Distribution of voiced variants according to sex and phrase position, /VsC_{voiced}/ context.*
4.4. Individual Variation

Finally, individual variation across the Mexican speakers in voicing assimilation of /s/ is considered. Table 5 presents the frequency of voicing of /s/ according to phonetic context for each individual speaker. Considering the voicing context, preceding a voiced consonant, considerable variability in the frequency of voicing is observed across the 12 speakers, from a tendency for voiceless sibilants (speaker 4M) to categorical voicing (speaker 3M).

Table 5. Percentage of /s/ voiced by phrase context for each speaker.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Voiced consonant</th>
<th></th>
<th>Vowel</th>
<th></th>
<th>Voiceless consonant</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>internal</td>
<td>final</td>
<td>total</td>
<td>internal</td>
<td>final</td>
<td>total</td>
</tr>
<tr>
<td>3M</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>14.3</td>
<td>30.0</td>
<td>23.5</td>
</tr>
<tr>
<td>12M</td>
<td>100.0</td>
<td>87.5</td>
<td>92.9</td>
<td>50.0</td>
<td>90.0</td>
<td>72.2</td>
</tr>
<tr>
<td>9F</td>
<td>100.0</td>
<td>77.8</td>
<td>86.7</td>
<td>12.5</td>
<td>20.0</td>
<td>16.7</td>
</tr>
<tr>
<td>8M</td>
<td>100.0</td>
<td>71.4</td>
<td>85.7</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>2F</td>
<td>100.0</td>
<td>28.6</td>
<td>61.5</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>14M</td>
<td>83.3</td>
<td>12.5</td>
<td>42.9</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>5F</td>
<td>83.3</td>
<td>.0</td>
<td>38.5</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>1F</td>
<td>80.0</td>
<td>66.7</td>
<td>71.4</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>10F</td>
<td>66.7</td>
<td>33.3</td>
<td>46.7</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>7M</td>
<td>33.3</td>
<td>87.5</td>
<td>64.3</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>6F</td>
<td>33.3</td>
<td>57.1</td>
<td>46.2</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>4M</td>
<td>33.3</td>
<td>12.5</td>
<td>21.4</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
</tr>
</tbody>
</table>

Voicing of the sibilant in the contexts where voicing is not expected (preceding a vowel and preceding a voiceless consonant) occurs only in the productions of 7 of the 12 speakers, and occurs at low frequencies (i.e., in less than one-third of the tokens)\(^\text{17}\). However, voicing in the unexpected context is found as a tendency for one speaker, 12M. This speaker tends to voice /s/ in intervocalic position, with voicing in 72\% of these contexts. Figure 6 presents an example of voicing of intervocalic /s/ by speaker 12M in the token José.\(^\text{18}\)

\(^{17}\) An anonymous reviewer pointed out that the three speakers who voiced /s/ in intervocalic position were also the three speakers with the greatest frequency of voicing in the predicted context (preceding a voiced consonant). There does not seem to be such a relation for those who voiced /s/ before a voiceless consonant.

\(^{18}\) We do not have a clear explanation for 12M’s patterns; however, we note that intervocalic voicing of /s/ has been recently noted in other dialects of Spanish (Torreira \textit{submitted}, Garcia 2011).
Figure 6. Example of phonemic /s/ realized as voiced [z] preceding a vowel in the name José.

4.5. Linear Mixed Model

To test whether the observed effects of the factors following segment, phrase position, and sex on the voicing of /s/ were statistically significant, a linear mixed model was run with the percentage voiced (the proportion of the /s/ segment that is voiced) as the dependent variable. The independent variables included in the model were following segment type, phrase position, and sex (fixed effects), as well as individual speaker and token (random effects). Main effects and interactions between the following segment, phrase position, and sex were tested, and a Bonferroni adjustment for multiple comparisons was applied for posthoc analysis.

A three-way interaction between following segment, phrase position, and sex was not found and was thus removed from the model. Main effects were found for following segment type (F(2, 23.594) = 124.111, p = .000) and phrase position (F(1, 23.607) = 9.013, p = .006), but not for sex. Also, significant interactions were observed for sex by following segment type (F(2, 518.586) = 6.840, p = .001), sex by phrase position (F(1, 518.111) = 7.718, p = .006), and phrase position by following segment type (F(2, 23.605) = 4.006, p = .032). A description of each significant main effect and interaction is provided in the following paragraphs.

The main effect of the following segment type revealed a significantly greater percentage voiced of /s/ when /s/ precedes a voiced consonant than when /s/ is found before a voiceless consonant or a vowel, p = .000 (Figure 7). However, the percentage voiced is not significantly different whether /s/ precedes a voiceless consonant or a vowel.
Figure 7. Average percentage voiced according to following segment.

Second, a main effect for phrase position reveals that the percentage voiced is significantly greater in phrase-internal position (with an average of voicing of 41% of the segment) than in phrase-final position (average of 35% of the segment is voiced), p = .006 (Figure 8).

Figure 8. Average percentage voiced according to phrase position.

The model also revealed three significant interactions between the variables. First, there was an interaction between the following segment type and the phrase position. The effect of phrase position is observed only for the context of /s/ preceding a voiced consonant. In this context only, there is a greater percentage voiced in phrase-internal position than in phrase-final position (Figure 9).
Although there is no main effect for sex, differences in voicing of /s/ are found for both male and female speakers when the interactions of sex by phrase position and sex by following segment type are considered. First, we find that the observed difference in voicing according to the position within the phrase is restricted to the female group only. The females have a significantly smaller percentage voiced in phrase-final position than in phrase-internal position. This significant difference according to phrase position does not occur with the male group (Figure 10).

**Figure 9.** Average percentage voiced according to following segment and phrase position.

**Figure 10.** Average percentage voiced according to sex and phrase position.
We also note a difference in voicing between males and females according to the following segment. There is a pattern observed in Figure 11, reflecting an interaction between sex and following segment type. While both males and females have a greater percentage voiced before a voiced consonant than before a vowel or voiceless consonant, there is a trend in which the percentage voiced preceding a voiceless consonant is larger than preceding a vowel for the females and the opposite trend for the males (greater voicing preceding a vowel than a voiceless consonant). The vowel-voiceless consonant difference is not significant for either group (females \( p = .072 \), males \( p = .184 \)), however we believe that a significant difference may emerge with a larger sample size (N=6 females and N=6 males).

**Figure 11.** Mean percentage voiced according to sex and following segment type.

### 4.6. Characteristics of the Voiced [z] and Voiceless [s] Variants

We now turn to a description of the voiced and voiceless variants of Spanish /s/ in terms of duration. The descriptive statistics for the total duration of the segment, the duration of voiced portion of the segment (voiced sibilance), and the calculated percentage voiced are displayed in Table 6 below. The descriptive statistics are presented separately for each phonetic context since the average duration of the /s/ segment is longer in intervocalic position than in coda position as reflected below. Duration of voicing reflects left-edge voicing (see Section 3.3).

<table>
<thead>
<tr>
<th></th>
<th>/NsC\textsubscript{voiceless}/*</th>
<th>/NsV/</th>
<th>/NsC\textsubscript{voiced}/*</th>
</tr>
</thead>
<tbody>
<tr>
<td>([s])</td>
<td>(N = 177)</td>
<td>(69.1 \text{ ms (19.92)})</td>
<td>(13.8 \text{ ms (6.39)})</td>
</tr>
<tr>
<td>([z])</td>
<td>(N = 10)</td>
<td>(58.2 \text{ ms (25.48)})</td>
<td>(47.8 \text{ ms (10.27)})</td>
</tr>
</tbody>
</table>

*Table 6.* Descriptive statistics of voiceless and voiced variants of /s/ for each phonetic context, N=567 tokens (SD).
First, we note in Table 6 that the sibilant in syllable-initial position (intervocalic /s/) is longer in duration than /s/ in coda position; this holds true for both the voiced [z] and voiceless [s] productions. Second, the mean duration of [z] is shorter than [s] for each of the contexts, and most notably in intervocalic position.

We also observe in Table 6 that even the voiceless variant [s] is characterized by some period of voiced sibilance. The average proportion of voicing is 16%, 21%, or 22% of the segment, for intervocalic, voiceless consonant, and voiced consonant contexts, respectively. Although the percentage voiced is different for the intervocalic context, the average duration of voiced sibilance is constant across all three contexts, with an average of 14-15 ms of initial voicing. Thus, what appears to be a slightly smaller percentage voiced for the intervocalic context is in fact a similar duration of initial voicing of the sibilant followed by a longer duration of voiceless sibilance of intervocalic /s/.

A linear mixed model was run to determine if the effects of voicing and of following segment type on the duration of the sibilant segment were significant. The dependent variable included in the model was total duration of /s/, with fixed effects of voicing of /s/ and following segment type, and random effects of speaker and token (lexical item).

Significant main effects on duration of the sibilant were found for the following segment type (F(2, 84.765) = 22.659, p =.000) and for voicing (F(1, 541.809) = 11.594, p = .001). Bonferroni post hoc tests reveal that /s/ is significantly longer in intervocalic position than when it precedes a voiced consonant or a voiceless consonant. For the main effect of voicing on sibilant duration, we find that [z] is significantly shorter in duration than [s].

An interaction between voicing and following segment type was found to be approaching significance (F(2, 539.617) = 2.699, p = .068). While a significant difference in duration of the segment was found between [z] and [s] in intervocalic position (p = .000) and preceding a voiced consonant (p = .001), a significant difference in duration of [z] and [s] preceding a voiceless consonant was not found. As there was variation in duration of [z] preceding a voiceless consonant and very few tokens (N=10), it may be that with a greater number of cases of [z] preceding a voiceless consonant a significant difference in duration may emerge as in the other contexts. This interaction is illustrated in Figure 12 below.

![Following Segment Type](Image)

Figure 12. Mean duration of /s/ (ms) according to voicing and following segment type.
5. Discussion of Findings

5.1. Voicing Assimilation Rule

The voicing of /s/ in Mexican Spanish among young adult speakers is not realized as a categorical process. The data followed a general pattern of voicing assimilation in the direction of the expected contexts, but there were phrase contexts that presented increased variation toward non-voicing where voicing was predicted. There was an absence of voicing of /s/ in the expected context (preceding a voiced consonant) in 37% of the data for all speakers. We also found cases of /s/ voicing in contexts were the allophone [z] is not supposed to occur, namely before a voiceless consonant and in intervocalic position. These observed cases of voicing of /s/ in the unexpected contexts in the Mexico City data support previous reports of voicing of /s/ preceding a voiceless consonant or vowel (Robinson 1979, Torreblanca 1978, Obaid 1973, Martín-Butragueño in press).

Our experiment also quantifies an expected, but previously unreported for Spanish, co-articulatory effect of vocalic preservative voicing into the sibilance, or frication, of the /s/. Our findings reveal that on average, tokens of the sibilant realized as voiceless were characterized by 14.4 ms of voicing co-occurring with frication at the beginning of the segment (SD = 6.89). This initial period of voicing comprised on average 21% of the /s/ segment duration for coda [s] and 16% of the segment duration for syllable-initial (intervocalic) [s]. Voicing that continues during the consonant constriction of voiceless obstruents is also reported in other languages, such as English (Stevens et al. 1992) and French (Darcy & Kugler 2007).

Concerning the nature of the assimilation process employed in /s/ voicing assimilation, our data reveals that there is a small but consistent degree of progressive assimilation of voicing from the vowel onto the consonant regardless of the postconsonantal segment. To some degree the voicing of /s/ should be viewed as both progressive and regressive as both segments (the preceding vowel and the following voiced consonant) play a role or contribute to the degree of /s/ voicing. We also saw a small number of tokens with progressive assimilation to the /s/ with a brief period of aperiodicity followed by anticipatory voicing for the following voiced consonant.

5.2. Segmental Duration

The present analysis found durational differences of Spanish /s/ according to the phonetic context and the variant (voiced or voiceless) produced. We find that the voiced sibilant is on average shorter in total duration than the voiceless sibilant in all three phonetic contexts (58 ms vs. 69 ms; 73 ms vs. 97 ms; and 60 ms vs. 70 ms; for the voiceless consonant, vowel, and voiced consonant contexts, respectively). This durational difference of obstruents according to voicing, with voiced obstruents being shorter in duration than their voiceless counterparts, has been reported previously in several other languages (e.g., Klatt 1976).

As the majority of the realizations of /s/ had an initial period of voiced sibilance, an initial examination of the duration difference may lead one to propose that what appears to be voicing of Spanish /s/ is in fact a simply a reflection of a shorter segment that ends before devoicing is activated. However, this is shown not to be the case, as the duration of the voiced sibilant, although shorter than the voiceless sibilant, is considerably longer than the average period of initial voicing in the voiceless cases. Also, although syllable-initial /s/ (in intervocalic position) is found to be longer in duration than coda /s/, the duration of initial voicing of the sibilant segment is similar across both contexts in the voiceless tokens, averaging 14-15ms in length. This initial period of voicing may not be surprising, however, and suggests a purely phonetic effect due to co-articulation. This co-articulation effect has been described by Stevens in the case of voiceless fricatives and intervocalic voiceless unaspirated stops, in which he describes a continuation of vocal fold vibration for two to three cycles after the stop closure (1998).

5.3. Phrase Position

Our findings indicate that the phonological process of voicing assimilation varies by phrase position. This variation in frequency of voicing reflects sociolinguistic variables in that it is only the female speakers that show a significant modification of voicing. Previous literature did not recognize
the role of phrase position in the application of the voicing process. A question that remains is whether this effect of phrase position is limited to Mexico City or if it occurs across other varieties of Spanish. One possibility is that this reflects a final devoicing, also found in vowel reduction between sibilants, /enton.ses/ → [εn.ton.s's]19, and in the assimilation of phrase-final /ɾ/ in this variety of Spanish (see Matus-Mendoza 2004). Since the application of the voicing process varies by phrase position, rule- or constraint-based analyses will need to account for phrase position to accurately model the current findings.

5.4. Sex of Speaker

As previously mentioned our data suggest differences in the application of voicing assimilation depending on speaker sex. This variation was not statistically significant, but general tendencies were observed. The male speakers presented a greater frequency of voicing of /s/ compared to the female speakers. There was also a difference between the two sexes in effect of phrase position, with females voicing less in phrase final position than phrase-internal position, but no such effect was found for the male speakers. There may also be additional sociolinguistic factors or interactions with other phonetic processes that contribute to the production patterns observed in the current corpus of university students. For example, it is possible to have social class distinctions among students of a public university such as the Universidad Autónoma Metropolitana-Iztapalapa. Current research on speech of young speakers from Mexico (Holguín-Mendoza 2011) reports stylistic variation regarding distinctions in social stereotypes commonly referred to as “naco” or “fresa” based on certain characteristics. These results show how young speakers, especially women from different socioeconomic groups who aim to perform identities of social distinctiveness (fresa), diverge in their use of several linguistic elements to differentiate themselves from the stereotyped speech of working classes (naco). We also note that the speakers that voiced the /s/ in intervocalic position where voicing is not expected were male, while interestingly, it was the females who produced the few voiced tokens of /s/ in the context preceding a voiceless consonant.

While our experimental contexts were not designed to study final devoicing (see above), our observation is that females tended to devoice finally more so than males. Indeed, this gender difference has also been reported for assimilation of word-final /ɾ/ in Mexican Spanish, with female speakers assimilating final /ɾ/ more than male speakers (Matus-Mendoza 2004). Future study should consider the potential relationship between final devoicing and sibilant voice assimilation in this dialect of Spanish. Future study might also reveal that the interaction of phrase position and sex (observed here in the Mexico City corpus) is limited to those dialects of Spanish characterized by final devoicing.

5.5. Individual Variation

Variation in the frequency of voicing of /s/ was found across individual speakers20. While several speakers voiced /s/ in the predicted voicing context in the majority of the tokens, other speakers showed a tendency in the opposite direction, with voicing of /s/ in less than half, and as few as 21%, of the voiced consonant conditions. Thus, individual variation is considerable in the application of the process. It should also be noted that both male and female speakers were found at both ends of the spectrum (tendency to voice or tendency not to voice). One motivation for individual variation lies in the lack of contrast between the two allophones. This high rate of individual variation also creates challenges for rule- or constraint-based explanations of voice assimilation.

19 In a study of Mexican intonation, final devoicing of final vowels and nasal vowel syllables were also reported in Pueblo Mexico Spanish, a near neighbor variety to the dialect studied (Willis 2005). Instances of final vowels that were devoiced were also found in the current Mexico City corpus.
20 Individual variation will often in part correspond to sociolinguistic variables that can be discovered and tracked with a large corpus of speech. With twelve informants in this study, which is relatively small by sociolinguistic standards, we are not able to identify a specific motivation for the increased use of this variant. However, studies of other phenomena (e.g., the production of the Spanish trill, Willis 2006) have encountered a similar occurrence of individual variation.
In addition to individual variation in the voicing of /s/ preceding a voiced consonant, variation was also observed with respect to the occurrence of the voiced sibilant in the contexts where [z] is not predicted to occur. Only 4 of the 12 speakers produced any cases of voicing preceding a voiceless consonant, and only 3 speakers realized voiced tokens of /s/ in intervocalic position. These cases of voicing of /s/ in the unexpected contexts tended to occur infrequently, in 1 to 3 tokens per speaker. However, there was one exception to this trend; speaker 12M displayed a tendency to voice /s/ in intervocalic position (in 72% of the contexts)\textsuperscript{21}. Potential linguistic and sociolinguistic factors relating to the occurrence of voicing of intervocalic /s/ across other speakers and dialects should be investigated in future research.

5.6. Limitations and Future Direction

The current study was limited to word-internal contexts and only addressed a portion of potential phonological contexts (see Section 3.2). The voicing assimilation rule is also claimed to occur across word boundaries and is not addressed in the current data set. Another potential variable is that of lexical stress. Campos-Astorkiza (2010) examines the role of stress and phonetic realization of the following consonant in the application of voicing. She finds no evidence for effect of stress on voicing, but does find a relationship between sibilant voicing and the phonetic realization of the following stop consonant.

Future studies could also address the role of lexical frequency on voice assimilation in Spanish. While this study was not designed to specifically examine the effect of lexical item or token on voicing, we did notice variation in frequency of voicing across lexical items in the data set. For example, certain items were almost always voiced, such as Osvaldo, beisbol, and bisnietos (more than 85% of the time), while other items tended not to be voiced, such as isla and esmalte (less than 40% of the time).

Finally, our findings are suggestive of the role of sociolinguistic variables, and past research on Mexican Spanish has found considerable variation within a specific geographic region based on these social variables. However, the role of sociolinguistic variables will require a larger population representative of various social levels to fully understand the current system.

6. Conclusions

Voicing assimilation of /s/, although indisputably present, is not a categorical process in Mexican Spanish. This variability in voicing assimilation was demonstrated through the systematic analysis of the allophonic realizations of the sibilant by a homogenous group of speakers from a /s/-conserving variety of Spanish.

Linguistic and extralinguistic factors were found to account for some of the variation observed in the voicing of /s/ in this variety of Spanish. The voicing of /s/ varied according to sex of the speaker and by position within the phrase, with female speakers voicing less frequently at the end of the phrase. This result suggests that there may be sociolinguistic factors relating to voicing assimilation of /s/, and that these factors should be explored in future work that considers extended social and regional groups.

We also document phonetic differences in the realization of the Spanish sibilant according to syllable position and to voicing. Spanish /s/ is longer in syllable-initial position (intervocalic /s/) than in syllable-final position. The voiced variant [z] is also shorter than the voiceless variant [s], across all contexts. Also, even in the voiceless variant [z], there is an initial period of voiced sibilance, extending from the voicing of the preceding vowel. This suggests that voicing assimilation of /s/ may, to some degree, be both progressive as well as regressive.

The results also demonstrate that voicing assimilation of /s/ is highly variable across individual speakers, even within the same regional variety of Spanish. Although all speakers exhibited voicing of

\textsuperscript{21} It is has been observed in other studies of a similar informant pool size that a single speaker may behave differently compared to the norm. For example, Willis (2006) reports one speaker out of 12 that consistently produced an assibilated trill in Santo Domingo, Dominican Spanish compared to the other participants.
preceding a voiced consonant to some degree, individuals varied in the frequency in which /s/ is voiced, from a tendency not to voice to near-categorical voicing. We also document active voicing assimilation of /s/ where it is not predicted to occur for one speaker in the corpus, in intervocalic position. Future work is required in order to determine to what extent voicing in unexpected contexts occurs across other Spanish-speaking individuals and varieties, and whether this is due to free variation or is suggestive of a change in progress, with voicing assimilation extending to occur before any voiced segment (consonant or vowel).

Appendix

List of tokens elicited in the contextualized picture description task

<table>
<thead>
<tr>
<th>/VsCvoiced/</th>
<th>/VsCvoiceless/</th>
<th>/VsV/</th>
</tr>
</thead>
<tbody>
<tr>
<td>chisme</td>
<td>discos</td>
<td>vasos</td>
</tr>
<tr>
<td>fantasma</td>
<td>mascaras</td>
<td>mesas</td>
</tr>
<tr>
<td>esmalte</td>
<td>moscas</td>
<td>nueces</td>
</tr>
<tr>
<td>Ismael (2)</td>
<td>pescados</td>
<td>casas</td>
</tr>
<tr>
<td>Esmeralda (2)</td>
<td>refrescos</td>
<td>osos</td>
</tr>
<tr>
<td>bisnietos</td>
<td>Oscar (2)</td>
<td>huesos</td>
</tr>
<tr>
<td>cisme</td>
<td>Francisco (2)</td>
<td>narices</td>
</tr>
<tr>
<td>durazno</td>
<td>revistas</td>
<td>lápices</td>
</tr>
<tr>
<td>asno</td>
<td>postres</td>
<td>payasos</td>
</tr>
<tr>
<td>beisbol</td>
<td>vestidos</td>
<td>Jose (2)</td>
</tr>
<tr>
<td>isla</td>
<td>Cristina (2)</td>
<td>Cesar (2)</td>
</tr>
<tr>
<td>Osvaldo (2)</td>
<td>Ernesto (2)</td>
<td>Luisa (2)</td>
</tr>
<tr>
<td>(rosbif)</td>
<td>espejo</td>
<td>Susana (2)</td>
</tr>
</tbody>
</table>

Table 7. Relevant tokens, word-internal context

References


