The Role of Structural Position in L2 Phonological Acquisition: Evidence from English Learners of Spanish as L2

Gabriela Vokic
Southern Methodist University

Abstract: In this pilot study, the speech of 12 adult native speakers of English with intermediate to intermediate-high proficiency in Spanish as a second language (L2) was analyzed to determine whether L2 learners rely on distributional information in the process of L2 speech learning and if so, if similar or dissimilar distributional patterns of sounds are more easily acquired. The parameter for (dis)similarity was set around the notion of structural position in combination with native language (L1) and L2 phonemic inventories. The results show that the subjects were consistently more successful in producing the phonemes with overlapping distributional patterns in L1 and L2 than phonemes whose distribution differed in L1 and L2 as well as novel L2 contrasts.

Key words: difficulty/ease of L2 acquisition, (dis)similarity, L2 phonological acquisition, L2 speech learning, structural position

Language: Spanish

Introduction
A topic that has received a lot of attention when discussing second language (L2) phonological acquisition is the role that similarity and dissimilarity of native language (L1) and L2 sound systems play in the process of acquisition of L2 phonology. This topic also represents one of the most controversial topics related to L2 phonological acquisition.

It is generally accepted that in the process of L2 acquisition, learners identify certain sounds in the L2 with L1 sounds. Researchers have descriptively referred to this tendency as a gravitational attraction (Schouten, 1975) and assimilatory pull (Best, McRoberts, & Sithole, 1988) that L1 sounds exert on L2 sounds. This phenomenon is also known as interlingual identification or equivalence classification, and it is an important cognitive mechanism involved in speech learning because it permits learners to make perceptual groupings of a wide variety of disparate phones with a common communicative function (Flege, 1992). An expected consequence of this process is that certain L2 sounds will be regarded by the learner as similar to L1 sounds, while others will be classified as different. Thus, equivalence clas-
sification inarguably exists, but researchers cannot agree on whether it facilitates or impedes the attainment of native-like pronunciation in the L2. Moreover, not only are there different opinions about which phonological phenomena, similar or dissimilar, are relatively easier to acquire, but there are two mutually exclusive opinions regarding this issue.

One position, predominantly associated with Lado (1937) and his contrastive analysis hypothesis, is that sounds similar in L1 and L2 are easier to acquire (Brière, 1966; Flege & Port, 1981; Johansson, 1973; Lado, 1957; Logan, Lively, & Pisoni, 1989; Politzer & Weiss, 1969; Trubetzkoy, 1939; Walz, 1979; Wode, 1976). Underlying the position that L2 learners will have more difficulty with new sounds is the assumption that the production of a sound requires the establishment of a central phonetic representation that contains information concerning both the perceptual target and a motor plan specifying how that target is to be achieved in speech production (Keele & Summers, 1976; Schmidt, 1976). Thus, according to this view, problems in pronunciation and foreign accent stem from both perception and production difficulties.

The opposing view, associated with Flege's speech learning model, is that sounds that differ in L1 and L2 will be more easily acquired by learners, whereas sounds similar in L1 and L2 will pose a problem for them (Flege, 1990; Valdman, 1976). It is predicted that late L2 learners will be less successful in learning similar L2 sounds precisely because they equate L2 sounds with sounds of the L1. Hence, similar sounds are difficult to acquire because speakers perceive and classify them as equivalents of existent sounds in L1 and find no reason to establish a new phonetic category. Conversely, new L2 sounds are easier to acquire because the differences are audible and readily noticed by learners. According to this view, difficulties in pronunciation and, in turn, foreign accent do not stem from production difficulties, but from perception difficulties.

Paradoxically, however, even though the role of similarity and dissimilarity of L1 and L2 sound systems in L2 phonological acquisition appears to be well studied and documented, what exactly constitutes categories such as similar and dissimilar seems to be a lot less well understood. First, as Young-Scholten (1985) points out, it is not clear what kind of similarity is being addressed—the typological one as defined by the linguist or the psychological one as defined by the learner—although the two should be one and the same if the linguists' claim that formal linguistics represents what the speaker knows about language is upheld.

More importantly, there is not one universally agreed upon definition or criterion as to what constitutes similar and what constitutes dissimilar in this context. The main reason for this situation is the very fact that these terms are not scientific and, as such, are difficult to define. As a consequence, the operationalization of the terms in question, if any, depends on each individual researcher. In this vein, researchers have been using different units of analysis as well as different levels of abstraction: phoneme inventories, distinctive features, acoustic cues, auditory properties, articulatory gestures, and in some cases even orthography, as well as native and nonnative-speaker intuitions (Major, 2002; Strange, 1999).

At any rate, whatever method is used to measure the degree of relative similarity or dissimilarity between L1 and L2 sounds, the underlying starting point always appears to be contrasting phonemic inventories of L1 and L2, and pointing out phonemic correspondences and oppositions across L1 and L2. The problem arises, however, when looking at phonemic correspondences or oppositions in L1 and L2 is no longer sufficient, because often a particular phoneme exists in the sound inventory of both languages, but its distributional patterns and surface realizations differ from one language to another. Therefore, it is the key theoretical assumption of this study that structural position and thus distributional properties of sounds must be taken into account when determining the
relative degree of similarity or dissimilarity between L1 and L2 sound systems.

Parameters for Analysis and Research Questions
This pilot study represents a first step in what will become a robust investigation of the role that structural position plays in the process of L2 phonological acquisition. The main goal of this pilot study is to experimentally test whether adult L2 learners acquire similar or different phonological phenomena more easily by setting the parameter for what constitutes similar and dissimilar between L1 and L2 phones around the notion of the structural position of a sound in a word, that is, whether a particular sound occurs word-initially, word-medially, or word-finally. For the purpose of this study, these structural positions are defined in the following way:

1. word-initial position = breath-group initial position (i.e., right after air inhalation),
2. word-medial position = word-internal syllable onset position preceded by a vowel, and
3. word-final position = breath-group final position (i.e., right before air exhalation).

Furthermore, the structural position parameter is combined with phonemic correspondences or oppositions across L1 and L2.

The main research questions this preliminary study aims to answer are as follows:

1. Do adult L2 learners rely on distributional information in the process of acquiring L2 phonology?
2. If so, does structural position have any bearing on what is considered by L2 learners to be similar to or dissimilar from their L1 sound system (and consequently, what is considered relatively more easy or more difficult) in the process of acquiring the L2 sound system?

To answer these questions, the speech of 12 adult native speakers of English acquiring Spanish as an L2 was analyzed.

Hypotheses
By combining structural position with phonemic correspondences and oppositions across L1 and L2, a gradual scale of difficulty in L2 phonological acquisition can be constructed:

Hypothesis #1: If L1 and L2 share an acoustically identical or similar phoneme that has the same phonotactic distribution in both L1 and L2, then this situation is regarded as identical and the sound in question will not represent a problem for the L2 learner. That sound will be positively transferred. (See Table 1.)

At this point, it is important to mention that the sound chosen to represent this category, that is /s/, displays a lot a variation cross-dialectally in the Spanish-speaking world. Namely, Spanish has three phonetically distinct varieties of this voiceless coronal fricative in the territories where it is spoken: the lamino-alveolar /s/, the apico-alveolar /s/, and the alveolar retroflex /∫/ (Hammond, 2001). Moreover, these three coronal phonemes have myriad alternate phonetic realizations in word-final and syllable-final environments (such as retention, voicing, deletion, and various degrees of aspiration). Even though the vast majority of native speakers of Spanish exhibit weakening of the /s/ in coda position in unaffected speech, it is assumed that those realizations of /s/ were not dominant in affected classroom speech and that the type of input the subjects in this study received was characterized by /s/-retention in a coda position. Additionally, given the dialectal varieties spoken by their instructors, the subjects were predominantly exposed to the lamino-alveolar variety of the Spanish /s/.

Given the fact that the phoneme /s/ in standard American English is also a voiceless lamino-alveolar fricative, for the purposes of this study the two sounds in question are assumed to be identical, because of the high level of their phonetic resemblance and because of their perfectly
overlapping distributional pattern in the two languages.

Hypothesis #2: If L1 and L2 share an acoustically identical or similar phoneme, but it has different phonotactic properties in L1 and L2, then this situation will be regarded as different by the learner and will represent a problem for acquisition. (See Table 2.)

What deserves attention at this point is the choice of the target phone /h/ as representative of this hypothesis for learners of Spanish as an L2. Namely, as suggested by the hypothesis, the only sound that qualifies to represent this parameter is a sound that is acoustically identical or similar in both English and Spanish and whose distributional patterns differ in that it is disallowed in a certain position in the native language but occurs in that very position in the L2. Based on the comparison of English and Spanish sound inventories and their distributional properties, the only sound that technically qualifies as representative of this category is /h/. The problems with this choice of target sound are various. First, not all dialects of Spanish have /h/ as a phoneme. In some dialects /x/ is the phoneme associated with graphic repre-
sentations "j," "ge," and "gi." Second, in dialects that do have /h/ at the phonemic level, this sound does not occur frequently in a word-final position. Moreover, only a handful of Spanish words end in /h/, and the only relatively frequently occurring one is relaj [re.loh] [watch] (Resnick, 1981). Furthermore, because of the strong preference of Spanish for open syllables, this sound is more often than not eliminated in almost all speech, particularly in casual speech. However, even though due to its instability word-final /h/ in Spanish is far from being an ideal candidate to test this hypothesis, it will be incorporated for analysis given the fact that classroom speech, especially at beginning levels, is far from casual and leans more toward the affected extreme, so there is a possibility that in the input the learners received they heard this sound in a word-final position.

Hypothesis #3: If the L2 has a phoneme that appears neither at the phonemic nor at the allophonic level in the L1, this will be regarded as different by the learner and will represent a problem in the acquisition of the L2 sound system. (See Table 3.)

The case described above is a major stumbling block in current theories of L2 phonological acquisition, because it has been posited to pose no problem for acquisition (e.g., Flege's speech learning model) as well as to represent the maximum level of difficulty for the learner (e.g., Lado's contrastive analysis hypothesis). The standpoint taken in this study is that, although in isolation L2 learners may be able to produce novel phonemic contrasts in a native-like fashion, maintaining that contrast in an equally native-like fashion in all L2 structural positions will present difficulties.

### The Role of Structural Position in L2 Phonological Acquisition

It is widely accepted that learning L2 speech constitutes considerably more than simply learning new pronunciation habits (e.g., James, 1988). But what exactly does it constitute?

If it is assumed that learning L2 speech is at least in certain aspects similar to acquiring L1 speech, then learning L2 phonology resembles learning L1 phonology in that the learners do not only need to establish phonetic categories, but must also discern the corresponding phonemic categories and acquire systematic knowledge of what patterns of sound sequences and other distributional patterns are permissible in a language if they intend to achieve native-like or near-native-like pronunciation in L2.

In determining the relative ease or difficulty of acquisition of L2 sound systems, considering L1 and L2 sounds in isolation (i.e., without regard to their phonological status in the L1 and L2 and without considering their distributional properties) can account for many sound substitutions observed in L2 speech. However, the same method falls short of accounting for cases such as the one observed in English pro-

---

**TABLE 3**

<table>
<thead>
<tr>
<th>Hypothesis #3</th>
<th>Language</th>
<th>Phone</th>
<th>W-I</th>
<th>W-M</th>
<th>W-F</th>
<th>Stimulus Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 = English</td>
<td>---</td>
<td>✓*</td>
<td>✓*</td>
<td>✓</td>
<td>rata [rá.ta] 'rat'</td>
<td></td>
</tr>
<tr>
<td>L2 = Spanish</td>
<td>/ɾ/</td>
<td>✓*</td>
<td>✓*</td>
<td>✓</td>
<td>carro [ká.ro] 'car'</td>
<td></td>
</tr>
</tbody>
</table>
nunciation of native speakers of Spanish. Namely, both English and Spanish sound inventories have the phone [m], which functions as a phoneme and has the same acoustic properties in both languages. Therefore, under both contrastive analysis hypothesis and the speech learning model, this sound qualifies as identical. More precisely, it qualifies as a sound that under no circumstances should pose a problem for acquisition. Thus, according to both views this sound simply should be positively transferred. Yet in casual speech, native speakers of Spanish display difficulties pronouncing a word-final English [m]. As a result, an English word such as album is pronounced [albun]—with the alveolar nasal [n] replacing what in English is a bilabial nasal [m].\(^8\) This phenomenon can be readily accounted for by the fact that, although [m] does exist in both Spanish and English, its distribution differs in these two languages. Specifically, in English this sound can occur in all three structural positions, whereas in Spanish it is banned from the word-final position. Instead, Spanish speakers use in this position a sound that is allowed in this position in Spanish, which is [n]. Thus, this example would seem to indicate that in order to achieve native-like pronunciation, an L2 learner must also acquire this type of distributional knowledge.

As far as phonotactics are concerned, L1 research indicates that phonotactic constraints are a distinct component of phonological knowledge and that infants, children, and even adults have a predisposition to extract these patterns from speech (Mattys & Jusczyk, 2001; Onishi, Chambers, & Fisher, 2002; Vitevich & Luce, 1998). Thus, it seems reasonable to assume that phonotactics is learned concomitantly with other components of the speakers’ L1 grammar. Bearing in mind that in the process of adult L2 phonological acquisition, the L1 definitely exerts an influence on the L2, it automatically follows that this type of learning, that is, learning of distributional patterns, will be affected by interference from the L1 as well. Indeed, some studies have suggested that L1 phonotactic constraints are carried over to the L2 and that such transfer influences the pronunciation of L2 sounds (Cebrian, 2004; Delattre, 1965; Rochet & Putnam Rochet, 1999; Solé, 1989; Tarone, 1987; Trammell, 1993). For instance, Solé observed that native English speakers’ pronunciation [no#se?] of the Spanish sentence No sé [no#se] [I don’t know] is an illustration of the operation of the English syllable structure constraint that prevents lax vowels from occurring in stressed open syllables. When this constraint is not present, as in the word él [el] [he], English speakers correctly produce a lax vowel. Thus, it was concluded that phonotactic constraints act as a filter of the acoustic signal and subsequent articulatory gestures.

Some other studies have also shown that perceptual difficulty of a novel L2 phonemic contrast may vary according to structural position (James, 1988; Major, 1986; Pisoni & Lively, 1995; Sheldon & Strange, 1982; Wieden, 1990). For instance, Major found that native speakers of English were producing the Spanish tap [r] more accurately in a word-medial position, as a result of the fact that the same phone appears in the same position (although as an allophone) in English. Additionally, Major's subjects were unsuccessful in producing the Spanish trill [r] in all three positions (i.e., word-initially, medially, or finally\(^5\)), but the few that were able to pronounce [r] were producing this sound successfully in particular in a word-medial position.

Evidence from L1 speech learning also supports the relationship between consonant acquisition and structural position, and demonstrates that some sounds are acquired first in one or two positions, but much later in another position. For example, stops are shown to be acquired later in word-final position than in other word positions (Templin, 1957). Longitudinal data from individual children have shown that some children acquire all segments first in one position: The children in Branigan's study (1976) acquired all new consonants
first in word-initial position, while the child described by Kiparsky and Menn (1977) produced all new segments first in word-final position. In addition, it has been shown that certain classes of sounds, such as velars and fricatives, may appear first in word-final or word-medial position (Edwards, 1979; Ingram, 1978). Although this tendency has not been supported by a large group of children (Stoel-Gammon, 1985), it may be reflective of individual variation.

Moreover, there is evidence from the L1 that acquisition of sounds is determined by a sound's functional load in the phonological system. For instance, opposing the claim that word-initial [v] is not acquired by English children until quite late, Ingram (1988) presented evidence that children acquiring Swedish, Estonian, and Bulgarian acquired this sound as one of the earliest due to the prominent role that word-initial [v] plays in these languages.

Additionally, it is a well-known fact in phonological theory that different structural positions can derive different surface realizations of the same phoneme. In that vein, various studies have attested to the tendency in adult L2 learners to focus attention only on those aspects of sounds needed for phonemic contrast (e.g., Best et al., 1988; Boomershine, Hall, Hume, & Johnson, 2008; Polka & Werker, 1994; Trubetzkoy, 1939). For instance, Best and colleagues report that native English speakers were able to discriminate the Zulu click contrasts and that performance on the most difficult contrast was not diminished even when the most obvious difference was eliminated.

However, this phoneme learning task may be complicated in instances where a phonemic category is implemented by more than one phonetic category (Flege, 1992). A concrete example can be found in English and Spanish. Both languages possess the sounds [d] and [ð], the voiced alveolar stop and voiced interdental fricative, respectively. However, these phones have a different status in the languages in question. In English those sounds are phonemes, whereas in Spanish [ð] is the allophone of the stop [d].

For Lado (1957), this kind of instance represents “a maximum learning difficulty” (Lado, p.15). In Flege's (1995) speech learning model, the structural position does not play a major role. This model does indeed posit that sounds in L1 and L2 are related perceptually to one another at a “position-sensitive allophonic level” (Flege, p. 239), but for Flege, this is synonymous with phonetic processing as opposed to more abstract phonemic processing.

Only a few studies have explicitly tested how a different phonological status of the sound affects L2 speech learning, or more precisely, whether allophony plays any kind of role in that process. For instance, Boomershine et al. (2008) directly tested the influence of the allophonic/phonemic distinction on perception. More precisely, they examined the ability of Spanish and English listeners to perceive phonetically identical sounds that have different phonological statuses in the two languages. Consistent with earlier studies, they concluded that contrast influences listeners' speech perception, whereas allophony does not.

In conclusion, as can be seen, structural position has an important impact on L2 speech learning in at least three ways:

1. Overlapping distributional patterns of sounds in L1 and L2 (or lack thereof) could have an effect on how L2 sounds are perceived and, in turn, produced;
2. Sounds in some structural positions may be relatively easier to acquire than in other positions; and
3. Contextually derived allophones may have an impact on relative ease or difficulty of L2 speech acquisition.

The present study will deal in more depth with the first point.

**Methodology**

**Participants**

This study had 12 participants, all naive adult native speakers of a midwestern variety of American English. The participants were all undergraduate university students.
of fifth- and sixth-semester intermediate Spanish classes at a large midwestern university campus. Their Spanish proficiency level was intermediate to intermediate high.\textsuperscript{13}

The participants first were screened by using a questionnaire to identify learners with the same or similar language learning experiences by monitoring the following variables: total number of years of L2 instruction, motivation, study abroad, travel experiences, heritage origin, and competence in other foreign languages.

All subjects reported a college grade point average index of 3.0 or higher. They also reported never having received a grade lower than the equivalent of A or B in any Spanish courses throughout their schooling. Furthermore, none of them were heritage speakers of Spanish and none had ever left the United States, either for study abroad, tourism, or any other purposes. Finally, all English-native subjects reported no competence in any second language other than Spanish.

The subjects in this study differ only in terms of their gender, age, the age when their L2 instruction started, and the total number of years of L2 instruction. There were five male and seven female subjects in this group. Their mean age was 20 at the time of recording. On average, their Spanish instruction had started at the age of 13.6. They had studied Spanish for an average of 6.1 years. The characteristics of English-native subjects are summarized in Table 4. While enrolled in the university, the input these learners were exposed to varied from different varieties of Latin American Spanish to some very limited nonnative Spanish. Finally, all the subjects in this study received compensation of $20 for their participation after the recording session. The specific purpose of the study was unknown to the participants.

**Stimuli Design**

Only one sound was chosen to represent each hypothesis. The reason for including
only one sound representative of each category was to avoid the creation of an extensive list of stimuli that would wear out the subjects during the data-collection procedure, as well as to avoid the possible interaction of confounding variables. Needless to say, there are other sounds in both English and Spanish that can fit the aforementioned parameters besides the ones incorporated directly into this analysis. However, not all were taken into account when designing the stimuli for the aforementioned reasons and thus, not all the existent possibilities were analyzed.

Once a phone representative of each parameter was determined, words containing the target phone in a particular structural position were selected. Four main criteria were used in the process of word selection: (1) that the word follows the CVCV (consonant-vowel-consonant-vowel) segmental pattern, (2) that the word is disyllabic, (3) that it carries penultimate stress, and (4) that the target consonant is surrounded by a variety of vowels in order to eliminate the possibility that surrounding vowels have an effect on how consonantal phones are produced in certain structural positions.

There were three tokens per each parameter. This amounted to a total of 33 stimuli. To prevent the purpose of the study from being transparent to the subjects, the stimuli were reshuffled and for each stimulus one additional distracter was added.

Data-Collection Procedure
In order to replicate to the maximum extent possible the intonation and pace of relatively unaffected speech, both the stimuli and the distracters were placed into a generic carrier sentence—**Veo la palabra STIMULUS en la pantalla** [I see the word STIMULUS on the screen]—and presented to the participants on a computer screen in a .ppt format with a 5-second interval between each sentence.

The participants were instructed to read aloud and at a normal pace the sentences on the screen containing the stimuli. Each set of stimuli was repeated and recorded four consecutive times. In each set, the ordering of the stimuli was the same. The recording of the first set of repetitions was discarded due to possible anxiety effects. The three remaining repetition sets were incorporated for analysis.

**Equipment**
Data collection was carried out in a sound-proof booth using a Sennheiser cardioid microphone model e815s and a Sony DAT recorder model TCD-D7. All the samples were recorded onto digital audiotapes. The settings used on the DAT recorder to carry out the recordings were as follows: sampling frequency: 48 KHz; microphone sensitivity: high; recording mode: manual; recording level: 6–7.

**Data Analysis**
The collected data were transferred from the DAT recorder to the computer and then redigitized at a 22050 Hz sampling rate using version 4.3.02. of Praat software and saved as a .wav file. A perceptual analysis of the data obtained was then conducted by playing the data from Praat and listening to them using Sennheiser headphones. Furthermore, in less clear cases, the perceptual analysis was aided by spectrographic analysis by using the same version of Praat software with default values. Finally, a frequency analysis was conducted.

**Results and Discussion**

**Hypothesis #1**
The stimuli used to test this hypothesis and the sounds expected to be produced by the subjects for the Spanish phoneme /s/ are presented below:

- **sano** [sá.no] [healthy]
- **solo** [só.lo] [alone]
- **silla** [sí.ja] [chair]
- **casa** [ká.sa] [house]
- **mesa** [mé.sa] [table]
- **piso** [pi.so] [floor]
- **gatos** [gá.jos] [cats]
- **gafas** [gá.fas] [glasses]
- **come** [kó.mes] [you (Sg.) eat]
The sounds observed during data analysis, that is, the sounds actually produced by English subjects in these instances, were [s] and [z]. As expected, [θ] was not found in the data because the subjects had not been exposed to Peninsular varieties of Spanish.

In terms of frequency, the target sound [s] in this data set was observed at the average rates of 100%, 100%, and 96.3% in word-initial, word-medial, and word-final position, respectively, whereas the substitute sound [z] occurred at the rate of 3.7% and was found only in word-final position. As can be seen, there was a strong predominance of [s] in all positions and for all subjects except one. Namely, in the speech of E-F8, the phone [z] was observed only in word-final position 44% of the time, on average.

Thus, in the case of hypothesis #1, the relatively high occurrence rates of [s] in all three structural positions would seem to suggest that the results in this study did not encounter difficulties in pronouncing the target sound. One way to account for the relatively easy and relatively high frequency with which subjects in this study produced the target sound in question would be to posit that the same phonological status of /s/ in both English and Spanish and its completely overlapping distribution in those two languages made it relatively easy for these subjects to perceive and, in turn, produce the sound in question. However, it should also be pointed out that these results could be an outcome of the native speakers of English relying on L1 orthography (and not a manifestation of the subjects having acquired a Spanish phonotactic rule), because the sound [s] is primarily graphically represented with the letter "s" in standard American English.

Hypothesis #2
The stimuli used to test this hypothesis and the target sounds expected to surface are shown below:
- reloj [re.loh] [watch]
- pedicoj [pe.đi.koh] [one-legged jump]
- relej [re.leh] [rut]

The sounds observed were: [ɔ], [h], and [ɔ], which overall occurred at the frequency rates of 53.8%, 34.3%, and 11.9%, respectively.

A closer look reveals that five subjects (E-F1, E-F6, E-F7, E-M4, and E-M7) constantly failed to produce [h]. Instead, a phonetic zero, [ɔ], was observed 100% of the time. In the speech of E-F8 [ɔ] also predominated at 67%, but was not exclusive. Instead, this subject also alternated between [h] and a very short and hardly audible [₃]. In the speech of the remaining six subjects (E-F2, E-F3, E-F5, E-M2, E-M3, and E-M5), [h] predominated and was observed on average 69% of the time. These six subjects pronounced [₃] 17.6% of the time on average and [ɔ] only 13.4% of the time on average. Finally, subject E-M3 used both [h] and [₃] with equal frequency, i.e., at 44% for both, with only one isolated instance of [ɔ].

Overall, in the case of this hypothesis, phonetic zero [ɔ] was observed most frequently among native speakers of English, since it was encountered in their speech 53.8% of the time on average. This kind of result offers support for the claim that an acoustically identical or similar phoneme with a different distributional pattern in the L1 and L2 will be difficult to perceive and, in turn, produce in the process of L2 speech learning.

These results, however, may be heavily confounded by the fact that word-final [h] as a realization of the grapheme "j" is a virtually nonexistent category in Spanish (recall the discussion in the Hypotheses section), as well as the fact that word-final position is an unstable phonological position in all natural languages.

Hypothesis #3
The stimuli used to test this hypothesis and the sounds expected to surface in the speech of the subjects are shown below:
- carro [ka.ro] [car]
- burro [bu.ro] [donkey]
- perro [pe.ro] [dog]
- rata [ra.ta] [rat]
- rojo [ro.ho] [red]
- risa [ri.sa] [laughter]
The sounds observed in the data set were: \([r], [\check{r}], [\tilde{r}], [\check{I}], [\tilde{I}], [r^2], [I],\) and \([I].\)

As can be seen, a variety of sound substitutions were observed during data analysis in the speech of native speakers of English in this data set. This wide array of produced sound substitutions potentially points to the uncertainty on the subjects' part as to the adequate articulatory behavior necessary to produce the target sound. Another potential explanation for the variability in the output is the fact that standard American English has only one rhotic sound at the phonemic level, whereas standard Spanish has two.

As far as frequency is concerned, in word-initial position the sound that occurred with the highest frequency rate was \([\check{I}],\) which is observed overall in this data set 67.5% of the time on average. The expected target sound \([r]\) occupied second place in terms of frequency, but its frequency rate dropped considerably to 13.9% on average. All the other observed variants were relatively infrequent, since they occur in the speech of native speakers of English at a rate lower than 10% on average.

A similar outcome was observed in word-medial position, as well. Namely, in word-medial position the most frequently observed sound was \([\check{I}],\) which overall occurred at a rate of 41.7% on average. The next most frequently occurring sound in this position was \([r],\) which occurs in the speech of native speakers of English 24.5% of the time on average. The expected target sound \([r]\) occupied third place in terms of frequency, since it occurred in this data set at a rate of 22.8% on average, whereas the remaining sound substitutions occurred at a rate of less than 5% on average.

Additionally, although in relation to this hypothesis, the English \([\check{I}]\) predominated in the speech of these subjects both word-initially and word-medially, there was one pattern that stood out. Namely, \([r]\) had a higher rate of occurrence in word-medial position than in word-initial position (22.8% word-medially vs. 13.9% word-initially). Recall that this result is in line with Major's (1986) finding, according to which Spanish trills (as well as flaps) are easier to pronounce in a word-medial position than word-initially for native speakers of English. A plausible explanation for this could be that \([r]\) is orthographically marked word-medially in Spanish, while it is not word-initially.

Thus, the frequency results obtained for this hypothesis point to the fact that a pattern like this one is indeed problematic for the learners, since in the speech of native speakers of English the predominantly pronounced variant, both word-initially and word-medially, was the American English phoneme \([\check{I}]\) and not the new L2 sound \([r]\).

Notice yet again that although this result can be explained as phonemic substitution, it can also be interpreted as reliance on L1 orthography because the grapheme “r” in standard American English is always pronounced \([\check{I}].\) If this is truly a case of L2 learners relying on English orthography (rather than their inability to produce a new L2 contrast), then this type of result is not at all surprising because the negative role of L1 orthography in L2 phonological acquisition has been discussed in relation to native-like acquisition of phonemes, contextually conditioned allophones, and sound systems in general. For instance, Young-Scholten (1995) points out that premature exposure to orthographic input (at or near the initial stages of L2 phonological development) can be expected to impede progression to native phonological competence in the L2. She claims that when the L2 learner who is literate in the L1 confronts the graphemes of a new language at the start of acquisition, prior to the establishment of L2 phonology, he or she will be compelled to search for the phonological constituents these graphemes might represent. Since the L2 phonology is not yet established, the L2 learner who is literate in the L1 will be able to access only the L1 phonology.

It should also be noted at this point that, besides the target sound \([r],\) some
of the other variants produced by native speakers of English in this data set (e.g., 
[r], [ɾ], [ɹ]) are also relatively frequently observed in these positions in the speech of native Spanish speakers because the phoneme /r/ has a lot of allophonic variation in Spanish. Thus, it is plausible to assume that native speakers of English produced these sounds because they were exposed to them through their Spanish instructors, and that based exclusively on the production of these sounds, they most likely would not be identified as nonnative speakers. Consequently, it is also plausible to assume that these three sounds did not appear in the speech of native speakers of English as new creations particular to interlanguage phonology.

Also, it can be speculated that in the case of hypothesis #3, many of the sound substitutions displayed by the learners in this data set are phonetically grounded and represent an inclination toward articulatorily simpler sounds. For instance, the choice of [ɾ] instead of [r] represents an obvious case of opting for an articulatorily simpler sound, as it is relatively easier to produce a flap instead of a trill while producing voicing at the same time. Also, the choice of [ɾ] makes sense from the standpoint of aerodynamics of speech, since it is relatively easier to pronounce a trill without having to produce voicing at the same time. Finally, the fact that native speakers of English found it easier to pronounce the Spanish trill word-medially and were struggling with it in a word-initial position can be accounted for by the fact that a considerable amount of air pressure is needed to pronounce this sound word-initially, but less is needed to do so word-medially.

Finally, a cumulative comparison of frequency results obtained for all three hypotheses reveals a constant decrease in average occurrence rates of target sounds from hypothesis #1 to hypothesis #3, i.e., 98.8% on average for the target sound in all three positions for hypothesis #1, 34.3% for hypothesis #2, and 18.3% for the target sound in both positions for hypothesis #3. More specifically, the highest target sound occurrence rates are observed in the pattern that was predicted to be identical, while they drop considerably (to lower than 35% on average) as the patterns become labeled as different (see Table 5), thus indicating that for native speakers of English in this study, nonoverlapping distributional patterns of sounds were relatively more difficult to produce than the overlapping distributional pattern. Therefore, it can be concluded that an account based on the distributional properties of sounds correctly predicts the relative degree of difficulty a learner will encounter in the process of classroom L2 phonological acquisition and that the hypotheses are confirmed.

It should be mentioned at this point that the results obtained in this study are intended to represent tendencies rather than absolute certainties. In this vein, the fact that in virtually all cases among native speakers of English there were certain subjects that indeed were able to accurately pronounce the target sound in a determined structural position would appear to suggest that native-like acquisition of L2 phonemes and distributional patterns is possible in the

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Acquisition Status</th>
<th>Target Sound</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Identical</td>
<td>[s]</td>
<td>98.8</td>
</tr>
<tr>
<td>2</td>
<td>Different</td>
<td>[h]</td>
<td>34.3</td>
</tr>
<tr>
<td>3</td>
<td>Different</td>
<td>[ɾ]</td>
<td>18.3</td>
</tr>
</tbody>
</table>
long run (although a separate longitudinal experiment would be needed to determine that with certainty). However, the time that it will take the learner to get to that point—the point of learning L2 speech in a native-like fashion—will depend on a variety of factors. In the present study, those decisive factors affecting native-like acquisition of L2 speech, and in particular the acquisition of L2 phonotactic knowledge, have been shown to be: (1) the functional load of sounds being acquired in different structural positions, (2) reliance on orthography, (3) the phonological status of sounds being acquired, (4) the relative articulatory simplicity or complexity of target sounds, and (5) the propensity of certain structural positions to be acquired earlier.

Conclusion
This pilot study represents a first step in what will become a robust investigation of the role that the structural position plays in the process of L2 phonological acquisition. It examined whether the structural position of a sound in a word has any bearing on the way L2 speech is perceived and, in turn, produced. More precisely, this study attempted to determine whether different distributional patterns of sounds coexistent in L1 and L2 affect in any way what is considered by classroom L2 learners to be similar and/or different in the process of L2 phonological acquisition. By setting the parameter for what constitutes similar and dissimilar in this context around the notion of structural position and combining this variable with phonemic inventories, a scale of relative difficulty of L2 phonological acquisition was proposed. It was posited that phonemes coexistent in L1 and L2 with an overlapping distribution in both languages will be perceived as identical and will be the easiest to acquire, whereas the most difficult instance for acquisition will be the case in which a learner has to develop a new L2 contrast and maintain it equally successfully in all structural positions in which it occurs in the L2. The former and the latter instances would be the extremes of the proposed continuum, with varying degrees of difficulty in between.

To test this assumption, the speech of 12 native speakers of English acquiring Spanish as an L2 was analyzed. It was found that overall native speakers of English were more successful in producing target sounds with overlapping distributional patterns in the L1 and L2 than those target sounds whose distribution differed in the L1 and L2 or new L2 contrasts, thus confirming that classroom L2 learners do rely on distributional properties of sounds in the process of L2 speech learning. It also was confirmed that overlapping and nonoverlapping distributional patterns do have an effect on what is regarded by these learners to be similar to and different from their L1 and consequently, what is regarded to be relatively easier or more difficult to pronounce.

Additionally, factors that obstruct information provided by distributional patterns and consequently interfere with the native-like production of L2 segments also were discovered in the process of data analysis among these L2 learners (such as reliance on orthography, the functional load of the sound in a particular structural position, the propensity of certain structural positions to be acquired earlier, and the relative inherent articulatory complexity of the sounds being acquired).

All of this points to the fact that learning L2 speech does not involve only processing an acoustic signal at a phonetic level, nor is it a mere comparison of phonemes in the L1 and L2, as previously suggested. Rather, it is a complex process in which many very closely related variables are intertwined.

The findings of this preliminary study contribute to the existing knowledge of L2 phonology by addressing one of the key issues in the field—variability in the input and the way in which this variability is processed by learners in developing new categories. In particular, the findings suggest new and more detailed ways of computing similarities between the L1 and L2, and thus making more specific predictions about L2 phonological development.
Notes
1. The author recognizes that there are sound contrasts in languages other than English and Spanish that lend themselves better to test out the hypotheses at hand. However, the choice of these two languages was practical and had to do with subject availability on the campus where data were collected.

2. In this study, two sounds are considered to be acoustically identical or similar when they have the same binary position, i.e., when they have identical frequencies, duration, and overall sound energy.

3. Phonotactics refers to the set of constraints on the possible sequences of sounds within a word, morpheme, or syllable (Trask, 1996).

4. In this study, the notions problem/problematic and difficult/difficulty mean that the acquisition of a particular pattern will be either significantly delayed or not acquired at all.

5. In this study, the term L2 learner always refers to a literate L2 learner. Thus, the hypotheses, findings, and conclusions of this study may not apply to illiterate L2 learners whose L2 acquisition proceeds entirely orally.

6. These three phonemes never coexist in the same dialect of Spanish. Any given dialect of Spanish will have only one of these sounds at the phonemic level, with the complete exclusion of the other two.

7. For more information on the type of input the participants in this study were exposed to and input variability, refer to the Methodology section.

8. This is not to say that native Spanish speakers are incapable of pronouncing word-final [m] when speaking English. They most certainly are, especially when circumstances require it, for instance, in formal situations. However, in casual speech, at a rapid pace, their English pronunciation will be filtered by Spanish phonological rules.

9. However, in Spanish the trill /r/ is found in word-final position only in emphatic contexts.

10. In this study, however, contrast is defined differently: as a surface, phonetic contrast, not a phonemic contrast (Boomershine et al., 2008).

11. More precisely, the intervocalic structural position selects the fricative allophone in Spanish, whereas in English the stop and the fricative are contrastive, so they both occur intervocally.

12. When used in this study, the term naive in the phrase naive native speaker refers to a native speaker with no background in linguistics as a science.

13. No explicit method was used to determine the proficiency level of the subjects; their proficiency level was determined solely based on the level of coursework in which they were enrolled.

14. Target sounds expected to surface are represented in bold.

15. The behavior of rhotic sounds in Spanish and their allophonic variation is by far the most complex issue in Spanish phonology. Besides the articulatory descriptions provided in Appendix A, the author cannot address these sounds in more depth due to space constraints. For a more thorough account of Spanish rhotics and their allophonic variation, see Hammond (2001).

References


APPENDIX A

List of Transcription Symbols

[r] – (voiced) apico-alveolar simple vibrant nonlateral liquid (here, used to represent both a flap and a tap)
[ɾ] – voiceless apico-alveolar simple vibrant nonlateral liquid
[r] – voiceless apico-alveolar multiple vibrant nonlateral liquid
[ɾ] – (voiced) apico-alveolar assibilated nonlateral liquid
[ɭ] – voiceless apico-alveolar assibilated nonlateral liquid
[ɭ] – (voiced) apico-alveolar rounded retroflexed nonlateral liquid
[ɭ'] – voiceless apico-alveolar rounded retroflexed nonlateral liquid
[r'] – (voiced) apico-alveolar retroflexed flap
[ɭ's] – voiceless apico-alveolar fricative with a short prevoicing portion
[h] – voiceless glottal fricative
[ʰ] – voiceless glottal fricative, hardly audible
[ʔ] – phonetic zero (silence)
[θ] – voiceless interdental fricative
[ð] – voiced interdental fricative