ATLANTIS
Journal of the Spanish Association of Anglo-American Studies
43.2 (December 2021): 1-21
e-ISSN 1989-6840
DOI: http://doi.org/10.28914/Atlantis-2021-43.2.01

Syllable Structure Effects in Word Recognition by Spanish- and German-Speaking Second Language Learners of English

MARIA TERESA MARTINEZ-GARCIA
University of Utah, Asia Campus
maria.martinezgarcia@utah.edu

Previous findings in the literature point to the influence that speech perception has on word recognition. However, which specific aspects of the first (L1) and second language (L2) mapping play the most important role is still not fully understood. This study explores whether, and if so, how, L1-L2 syllable-structure differences affect word recognition. Spanish- and German-speaking English learners completed an AXB and a word-monitoring task in English that manipulated the presence of a vowel in words with /s/-initial consonant clusters—e.g., *especially* versus *specially*. The results show a clear effect of L1 on L2 learners' perception and word recognition, with the German group outperforming the Spanish one. These results indicate that the similarity in the syllable structure between English and German fosters positive transfer in both perception and word recognition despite the inexact segmental mapping.

Keywords: syllable structure; epenthetic vowels; L2 learners; word recognition; speech perception

. . .

Efectos de la estructura silábica en el reconocimiento de palabras por parte de aprendices de inglés cuya lengua nativa es el español o el alemán

Si bien se ha demostrado que la percepción del habla influye en el reconocimiento de palabras, todavía no existe una comprensión suficiente de los aspectos específicos de la relación entre la primera (L1) y la segunda lengua (L2) que desempeñan el papel más importante en este sentido. Este estudio explora si, y si es así, cómo, las diferencias de estructura silábica entre la L1 y la L2 afectan al reconocimiento de palabras. Los aprendices de inglés que hablan español o alemán como lengua nativa completaron una tarea de percepción AXB y una tarea de monitoreo de palabras en inglés manipulando la presencia de una vocal en palabras que empiezan por grupos consonánticos con /s/ (por ejemplo, *especially* versus *specially*). Los resultados muestran un claro efecto de la L1 en la percepción y el reconocimiento de palabras de los estudiantes de L2, con el grupo alemán obteniendo mejores resultados que el español. Estos resultados indican que la similitud en la estructura silábica entre el inglés y el alemán fomenta la transferencia positiva de este conocimiento tanto en la percepción como en el reconocimiento de palabras, a pesar de que los segmentos no sean exactamente los mismos en ambas lenguas.

Palabras clave: estructura silábica; vocales epentéticas; aprendices de L2; reconocimiento de palabras; percepción del habla

1. Introduction

For some adult second language (L2) learners, to master a new phonology seems to be a hurdle that is difficult to overcome, something not commonly reported among young and adolescent learners. This generational difference has been explained by a series of different factors (for a detailed discussion, see Strange 1995; Bohn and Munro 2007). However, the entrenchment of the native language (L1) after years of being exposed to and using it has been proposed as the principal reason for these difficulties when learning a language in later life. In other words, when an adult learner tries to acquire the L2 phonology, the new phonetic system interacts with the one already established and fixed through years of use for their L1, and it is this that influences the development of L2 accent (Flege 1995; Flege et al. 1995; Flege et al. 1999). The main difference, then, between children acquiring their L1 and adults acquiring an L2 is that the latter filter L2 sounds with reference to previously established L1 phonetic categories. However, as proficiency in and exposure to the L2 increases, adult learners are expected to master the perception and production of novel L2 phones—i.e., sounds without an L1 phonetic counterpart. In contrast, L2 sounds that have an L1 counterpart but are realized in a phonetically different manner are predicted to be challenging for most learners, even at advanced levels of proficiency. That is, L1-L2 phonemic differences are expected to cause perception problems (Best et al. 1988; Best 1995; Flege 1995; Best and Tyler 2007).

Importantly, these potential phonemic problems linked to learners' L1 are not limited to the acquisition of L2 segmental contrasts, but extend to its phonotactics—the way segments fit together in permissible combinations. For example, while both Spanish and English have /s/ and /t/ in their phonetic inventories, only English allows them to be combined in this order in syllable- and word-initial position—e.g., Spanish *estudiar* versus English *study*. In fact, research has found that the perception of an L2 can be affected by how segments can be combined in relation to word and syllable boundaries (Dehaene-Lambertz et al. 2000; Dupoux et al. 2001; Matthews and Brown 2004; Kabak and Idsardi 2007). Hence listeners' perception of words is not only predicted to be influenced by the segments involved, but also by the language-specific rules governing how they can be combined.

It is not clear, however, how L1 phonotactics impacts the recognition of L2 words in continuous speech, as most existing studies have focused on words in isolation (Dupoux et al. 2001; Kabak and Idsardi 2007). Furthermore, it remains to be determined whether L2 learners experience speech processing difficulties only when the difference is attested in the exact consonant combination or also in the sound-feature combination. For example, German allows sibilant-initial consonant clusters—e.g., *Stuhl* [ʃtuːl] "chair"—but not /s/-initial consonant clusters, while Spanish does not allow either option. However, German-speaking L2 learners of English have not been reported to epenthesize before /s/-initial consonant clusters, as Spanish-speaking L2 learners of English tend to do (Carlisle 1997, 1998). This raises the question of whether the differences between Spanish, German and English would trigger similar difficulties among proficiency-matched L2 learners of English. If it was found that only Spanish-speaking learners show perception difficulties, this would indicate that L2 learners' speech processing difficulties are linked to L1-L2 differences in phonotactics—namely, constraints on the possible sequencing of exact consonants in word-initial position. However, if both groups of learners patterned similarly, the difference would be attributable to differences in syllable structure—namely, constraints on the type of segments allowed in syllable onset position—which indeed better accounts for some misperception problems.

The aim of the current study is, thus, to investigate the effect of L1 phonotactics on the processing of the L2 speech signal by late L2 learners—those who started learning the language after puberty—focusing on /s/-initial consonant clusters in English by adult Spanish and German L2 learners of English. The article starts by reviewing the literature on how language-specific rules governing the combination of segments impact perception of the speech signal and continues with the presentation of two perception experiments—Experiment 1, a perceptual discrimination AXB task, and Experiment 2, a word-monitoring task. Finally, the article concludes with a discussion of the implications of the results of these experiments for understanding how L1 phonotactics affects L2 speech perception.

2. Effects of L1 Phonotactics on L2 Speech Production, Perception and Word Recognition

When L2 learners try to produce a sequence of sounds that are not permitted in their L1—that is, when the sequence violates their L1 phonotactics—foreign accent can be particularly noticeable. Common L2 repair mechanisms in the production of L1 phonotactic violations include consonant deletion, prothesis, metathesis and vowel epenthesis (Anderson 1987; Broselow and Finer 1991; Eckman and Iverson 1993; Carlisle 1997, 1998; Hancin-Bhatt and Bhatt 1998; Abrahamsson 1999; Davidson et al. 2004; Davidson 2006), with vowel epenthesis being the most commonly noted strategy (Carlisle 1997). For example, as described earlier, English, but not Spanish, allows /s/ and many consonants to be combined in syllable- and word-initial position, with Spanish showing a more restrictive inventory of consonant clusters in onset position. Thus, Spanish L2 learners of English commonly introduce an epenthetic vowel before /s/-initial consonant clusters in their oral productions, with a word like *school* being produced as [es'kul] instead of [skul] (Carlisle 1997, 1998; Abrahamsson 1999).

L1 phonotactic constraints, though, influence L2 speech perception as well as production. Using a cross-linguistic speech perception experiment, Emmanuel Dupoux et al. (1999) examined the transfer of L1 phonotactic constraints on the perception of unfamiliar words by native speakers of Japanese, which has a rather restrictive syllable structure allowing only V, VV, CVN and CVQ sequences, where Q is a geminate. Their perception results were compared with those of French speakers, whose L1 has a more permissive syllable structure. Participants were presented with a series of nonce words, ranging from *ebzo* to *ebuzo*, where the medial vowel was shortened in segments of five pitch periods, that is, from a full vowel to no vowel. Unlike French listeners, most of the Japanese heard an illusory vowel in the illegal consonant sequence even when there was no vocalic segment between the two consonants. These findings were taken as evidence that Japanese listeners were perceptually repairing an illegal sequence of sounds by hearing an inexistent vowel in contexts that violated their L1 phonotactics.

These results have been replicated in several publications. In a follow-up study using event-related potentials (ERPs), Ghislaine Dehaene-Lambertz et al. (2000) examined whether participants with the same language backgrounds as those in Dupoux et al. (1999) and using the same stimuli would elicit a mismatch negativity (MMN). Participants heard a series of four stimuli, with an interstimulus interval of 600 ms, which met one of two conditions, i.e., whether the last word was the same as the previous three or was the deviant stimulus—*ebuzo* - *ebuzo* - *ebuzo* - *ebzo*). An MMN would suggest that a listener could distinguish the stimuli containing an interconsonantal vowel from those not containing it. The results revealed that whereas for native French listeners, a clear MMN was detected in their brain responses, the effect was weaker or completely absent in the Japanese listeners' brain responses.

Further evidence of the perceptual illusion phenomenon in Japanese subjects was provided in a study by Dupoux et al. (2001) that explored whether it was related to a top-down lexical effect. The C1C2 stimuli used in their previous study could have activated lexical items that contain C1/u/C2 and thus elicited the attested perception of an epenthetic vowel due to potential lexical effects. In their 2001 study, they manipulated the lexical neighborhood of the nonwords by having half of the stimuli be nonwords whose [u]-epenthetic counterparts are real Japanese words—u-set; e.g., *bakro*-*bakuro* "exposure"-*bakaro*—and the other half having no [u]-epenthetic counterparts in Japanese—non u-set; e.g., *arsi*-*arasi* "storm"-*arusi*. Notice that in the u-set the word containing /u/ is a real Japanese word, which could lead to more misperceptions of epenthetic vowels, but not in the non u-set. Participants were presented with the stimuli and were asked to transcribe them and judge whether or not they were real Japanese words. The results showed that, independent of the existence of a real counterpart in Japanese, both types of stimuli elicited equal amounts of [u] perception. Thus, no lexical effect by itself could explain the results found in the 2001 study, yet it is not known how this misperception problem may affect word recognition.

John Matthews and Cythia Brown (2004) presented additional evidence for the effects of L1 phonotactics on the perception of illusory vowels. They compared the performance of Japanese and Thai L2 learners of English in terms of their perception of [k.t] versus [kVt] clusters, also using nonce words and a perceptual AX discrimination task. The sequence [kt] is not possible in either language, although [k] is a possible syllable coda in Thai, but not in Japanese. Their results showed that Thai listeners were better at discriminating words with a vowel from those without. These results were interpreted as evidence that the misperception of consonant clusters is possibly triggered by L1 syllable structure constraints rather than by constraints on the permissible sequences of segments in the L1. However, in this study the two groups were not proficiency matched, which could have led to differences in their performance.

Further extending this line of research, BariŞKabak and William J. Idsardi (2007) looked at the perception of illusory vowels in word-medial clusters by Korean L2 learners of English. The authors wanted to determine whether the perception of illusory vowels is best explained by the nonoccurrence of certain consonants in coda position—e.g., *[c], *[g])—or consonantal contact restrictions that prohibit the combination of certain heterosyllabic consonants—e.g., *[k.m], *[l.n]. The results of their AX discrimination paradigm suggested that L1 syllable structure constraints rather than consonantal contact restrictions are responsible for these L2 English learners' perception of epenthetic vowels.

This phenomenon has not only been explored with Japanese and Korean L2 learners of English; two recent studies have examined the effect of misperceiving an epenthetic vowel in production, perception and word recognition of /s/-initial consonant clusters by German- and Spanish-speaking L2 learners of English (Martinez-Garcia and Tremblay 2015, n.p.). By means of the same word-monitoring task as is used in the present study together with a production task, we explored the relationship between both misproduction and problems in aural recognition of words with epenthetic vowels in /s/-initial consonant clusters. The German group, whose L1 allows for similar consonant clusters to appear in word-initial position, outperformed the Spanish group. The study also showed that only those Spanish speakers who recognized the word were less likely to mispronounce it. That is, the results indicated that L2 word-production difficulties might be closely tied to L2 word-recognition difficulties. In a more recent study, I tried to find an explanation for these word-recognition difficulties by linking them to perception problems (Martinez-Garcia 2018). Native speakers of Spanish at an intermediate to advanced level of proficiency completed both an AXB and a word-monitoring task where the presence of the epenthetic vowel was manipulated. The native speakers of Spanish in this study also struggled to perceive the presence or absence of the epenthetic vowel in both tasks, indicating that speech perception, in this case, was crucial for word recognition.

However, two main questions remain unanswered. The first relates to the potential effect of L1 phonotactics on the recognition of L2 words in continuous speech, not just in isolated words. If epenthetic vowels are perceived in isolated words, they should also be perceived in continuous speech. In fact, misperception in speech could potentially activate competitor words, making lexical access less efficient and L2 speech processing more difficult (Weber and Cutler 2004; Escudero 2007; Broersma and Cutler 2011; Martinez-Garcia 2018). The effects of this activation are, then, more likely to emerge in continuous speech rather than in isolated words. The second question is whether it is the exact consonant combination or the sound-feature combination that determines whether L2 learners experience difficulty. It should be remembered that, unlike in English, while neither Spanish nor German allows /s/-initial consonant clusters, German does allow sibilant-initial ones—e.g., *Straße* [ˈʃtʁaːsə] "street". Phonotactic effects would predict that both German and Spanish listeners would hear epenthetic vowels in English /s/-initial consonant clusters—in words such as *special*—as neither language allows this combination in word-initial position. In contrast, more abstract syllable-structure effects—constraints on the *types* of consonants allowed in syllable-initial position, e.g., sibilant-initial clusters rather than /s/-initial clusters—would predict that only Spanish listeners would hear epenthetic vowels. Thus, the current study addresses the effect of L1 phonotactics on the recognition of L2 words in continuous speech and the role of the exact consonant combination on L2 learners' perception difficulties.

3. Experiment 1: AXB Task

3.1. Participants

32 native speakers of English—16 females; mean age 23 years—32 Spanish-speaking—12 females; mean age 24 years—and 32 German-speaking learners of English—14 females; mean age 23 years—participated in this study. The native speakers were students at a Midwestern university in the US. Some of the L2 learners—20 Spanish and 12 German speakers—were tested at the same institution in the US, where they were undergraduates, graduate students or lecturers in the Spanish and German departments. The remaining participants were tested at the universities in Spain and Germany where they were students. The L2 learner groups, including all participants tested, were matched in terms of age of L2 acquisition—$t < |1|$—months of residence—$t(62) = -1.77$, $p > 0.1$—years of instructions—$t < |1|$—and proficiency—$t(62) = -1.12$, $p > 0.1$)—as established by a cloze test (Brown 1980). Learners' background information and proficiency scores are presented in table 1.

Table 1. L2 learners' background information

Language		Age of L2 acquisition	Years of instruction	Months of residence	Proficiency (/50)
Spanish L2 learners (32)	Average (SD)	12 (2.3)	12 (3.6)	15 (36.8)	25.94 (12.04)
	Range	9-17	6-22	0-156	6-48
German L2 learners (32)	Average (SD)	12 (2.3)	10 (3.9)	20 (38)	28.84 (9.3)
	Range	10-17	3-22	0-156	6-45

3.2. Materials

Participants completed an AXB discrimination task. 16 nonce words that began with an /s/-initial consonant cluster were created for this task in order to avoid participants using lexical information when judging the presence or absence of an epenthetic vowel. Each stimulus pair contained either a schwa—e.g., /əs.'lɛn/—or did not—e.g., /slɛn/—before the consonant cluster. The same materials were used in my 2018 study. The study also included 32 fillers. These nonce word fillers met one of two conditions: the first involved stimulus pairs that contained either /b/ or /v/—e.g., /'mi.bɹɛz/ versus /'mi.vɹɛz/—a contrast that is difficult for Spanish L2 learners of English as /v/ is not phonemic in their L1 but /b/ is. They included stimuli that differed in the presence of one phoneme—e.g., /snun/ versus /snu/. To make sure participants were paying close attention to the complete word, the sound contrasts of these two filler conditions appeared in different positions in the word—initial, medial or final.

The resulting words—stimuli and fillers—were checked by a native English speaker to confirm that they followed the phonotactic rules of English and did not resemble any real English word. To ensure that participants would not rely on the physical—acoustic—properties of the stimuli to do the task, three different speakers were recorded uttering them.

3.3. Procedure

The stimuli were presented by means of the stimulus-presentation software *Paradigm* (Tagliaferri 2005) and the same procedure was followed as in my 2018 study. Participants were instructed to listen carefully to a series of three nonce words and to choose whether the second word (X) was more similar to the first or to the third (A or B), with the interstimulus interval set at 1,000 ms. Participants made their decision and pressed one of the two buttons of the mouse—left for A, right for B. The next trial started immediately after the participants entered their response. A practice session of six stimuli with feedback—correct or incorrect response automatically coded by the program—preceded the main session of the experiment, which did not provide any feedback. All trials were randomized across participants.

The stimuli were always presented in the same order, with A being produced by Speaker 1 (Midwest dialect), X by Speaker 2 (East coast dialect) and B by Speaker 3 (Midwest dialect). The test items were presented in a Latin square design so that participants would hear either A or B as X but not both.

3.4. Data Analysis

The accuracy of participants was analyzed with a logistic regression model (Baayen 2008), using the *glm* package (Hothorn and Everitt 2014) in *R* (R Development Core Team 2009). L1 was considered as a categorical predictor with three levels—English versus Spanish versus German—the English group representing the baseline to which the other groups were compared. The effect of the predictor was assessed by means of log-likelihood tests comparing models with and without the predictor. Two sets of models were run—one on all the accuracy rates using L1 as predictor, and the other on the L2 learners' accuracy rates using L1, proficiency and their interaction as predictors. The effect of L2 proficiency was assessed by comparing models that included proficiency with models that did not; in each case, the model with the best fit was kept. Since the inclusion of proficiency did not improve the L2 model, only the analysis of all accuracy data is reported here. Participant and item were included as random variables.

3.5. Results

Figure 1 shows the mean accuracy results for the three groups, and table 2 the results of the logit regression model for all participants' accuracy.

Figure 1. Mean accuracy (standard errors) of the three groups in the AXB task

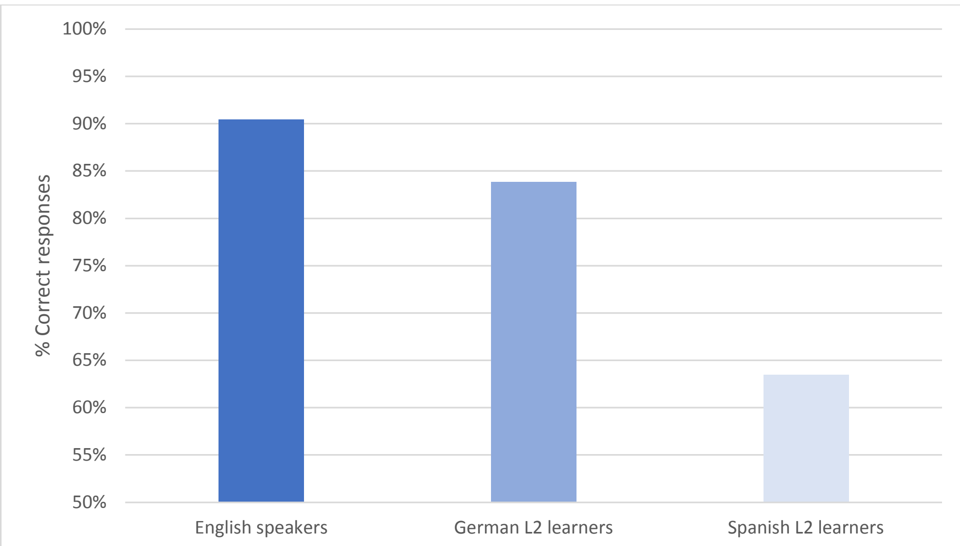


Table 2. Logit regression model on all participants' accuracy results

Variable	Estimate (SE)	z	p
(Intercept)	2.24 (0.15)	14.95	< 0.001
L1: German L2 learners	-0.03 (0.2)	-1.67	> 0.1
L1: Spanish L2 learners	-1.77 (0.2)	-10.07	< 0.001

The model summarized in table 2 reveals that while the German L2 learners of English performed in a similar way to native English speakers, Spanish L2 learners of English were statistically less accurate than native English speakers in discriminating whether stimuli with /s/-initial clusters contained a word-initial schwa or not. A follow-up analysis was run comparing the L2 learners' results to confirm that they patterned differently. The results showed a significant difference in accuracy rates between the L2 groups—$F(63) = -1.53$, $p < 0.01$.

3.6. Discussion

In this first experiment, an AXB task was used to examine whether proficiency-matched Spanish and German L2 learners of English would have similar difficulties in the perception of /s/-initial clusters in English. The results showed that the German learners, like native English speakers, could easily distinguish between /s/-initial cluster forms preceded by an epenthetic vowel and those that were not. In contrast, native speakers of Spanish showed a less accurate perception of this distinction. The results of the Spanish group matched the results reported in my 2018 study.

Two main conclusions can be drawn from this set of results together with those I reported in 2018. First, the Spanish speakers' results cannot be attributed solely to the fact that they were L2 learners of English, since proficiency-matched German learners performed as accurately as native speakers on the task. Second, the results of this discrimination task for the three groups were in accordance with the more abstract syllable structure constraints of each L1. The Spanish speakers' perception of the stimuli was shaped to conform to the syllable structure constraints of their L1, in which /s/-initial—and all sibilant-initial—consonant clusters are illegal. This perception problem mirrors the production errors that are commonly associated with Spanish L2 learners of English (Carlisle 1997, 1998; Abrahamsson 1999). In contrast, German L2 learners of English did not show major difficulties in perceiving the difference between the stimuli with and those without an epenthetic vowel, probably because German allows other sibilant-initial clusters even though it does not allow /s/-initial ones.

The results of this discrimination task add support to the body of evidence that suggests that it is syllable structure constraints rather than phonotactics that influence L2 learners' speech perception (Matthews and Brown 2004; Kabak and Idsardi 2007; Martinez-Garcia 2018). Here, L1 phonotactic effects predicted no difference between the two groups of learners, whereas more abstract L1 syllable structure constraints, which stipulate that only some types of segments are permissible in syllable and word onset position, predicted the pattern of results obtained. However, it remains to be determined how these perception errors affect L2 word recognition. From the literature, we know that difficulty in correctly perceiving L2 sounds can increase lexical competition and make word recognition less efficient (Weber and Cutler 2004; Escudero 2007; Broersma and

Cutler 2011; Martinez-Garcia 2018). Experiment 2 was thus created to examine how the perception of illusory vowels might impact L2 word recognition.

4. Experiment 2: Word-Monitoring Task

4.1. Participants

The same participants who completed Experiment 1 also took part in Experiment 2. In facts, all participants completed the word-monitoring task—Experiment 2—before the AXB task—Experiment 1—to ensure that their word recognition would not be influenced by what they might have explicitly noticed in the AXB task.

4.2. Materials

For this experiment, all groups completed a word-monitoring task with real words requiring them to keep track of a predesignated written target word in the acoustic input (Kilborn and Moss 1996). In this case, participants were given as target word one item from a minimal/near-minimal pair that either had or did not have a schwa in word initial position—i.e., *estate* versus *state*. One word of the pair was embedded in a semantically ambiguous sentence where either word was legitimate. Participants had to decide whether the target word was present in the sentence or not—i.e., match or mismatch the target word. One of the reasons to use this offline task is that it makes it possible to measure the participants' unconscious response to the presence of a potential epenthetic vowel. While the task puts a burden on participants' working memory, it is also a fairly direct measure of their mental representations (Kilborn and Moss 1996). Table 3 illustrates these four conditions with an example.

TABLE 3. Example of stimuli used in the word-monitoring task

Condition	**Written target**	**Auditory stimulus**
Match Stimulus with vowel	ESTATE	*I have lived in this <u>estate</u> for a long tim*e.
Match Stimulus without vowel	STATE	*I have lived in this <u>state</u> for a long tim*e.
Mismatch Stimulus with vowel	STATE	*I have lived in this <u>estate</u> for a long time.*
Mismatch Stimulus without vowel	ESTATE	*I have lived in this <u>state</u> for a long time.*

48 pairs—36 minimal and 12 near-minimal—manipulating the presence or absence of a vowel at the beginning of /s/-initial clusters were used—e.g., *state* versus *estate*. The near-minimal pairs were created by including inflected and derived forms of the words—e.g., both *estate* versus *state* and *estates* versus *states*. Semantically ambiguous sentences, checked by two native English speakers to ensure their plausibility and ambiguity, were created to prevent participants from using lexical cues when making their decisions. The location of the target word within the sentence was also manipulated to prevent participants from creating expectations, and the words never appeared after a word-final schwa.

Also included were 96 fillers that followed the characteristics of those described in Experiment 1—that is, minimal embedding in semantically ambiguous sentences. In other words, half the filler items differed in whether they contained /b/ or /v/—e.g., *bail* versus *veil*. The fillers were not only made up of minimal pairs; near-minimal pairs were also included to match, as closely as possible, all the characteristics of the target stimuli described earlier. The other half of the filler items differed in the number of phonemes they contained or in one of their phonemes—e.g., *stop* versus *top* or *veil* versus *bail*. The sound contrasts of these two filler conditions appeared in different positions in the word—initial, medial or final.

All sentences—experimental ones and fillers—were recorded by a female native speaker of American English with a Midwest accent—Speaker 3 from Experiment 1. A randomized Latin square design was used to present the stimuli to participants.

4.3. Procedure

As in Experiment 1, stimuli were presented by means of *Paradigm* (Tagliaferri 2005). In each trial, participants saw the target word in the middle of the screen in capital letters—e.g., ESTATE—for 1,000 ms. The word disappeared at the same time as the audio started playing a sentence that either included or did not include it—e.g., *I have lived in this estate for a long time* versus *I have lived in this state for a long time*. Participants had to click left on the mouse—labelled “yes”—if the sentence contained the word they had seen on the screen, or click right—labelled “no”—if it did not. Participants could respond as soon as they identified the word in the sentence—i.e., before the audio finished—or wait to the end of the sentence. Either way, as soon as they entered their response, the next trial started. The experiment started with six practice trials, which included feedback—correct or incorrect response automatically coded by the program. During the main session of the experiment, participants did not receive any feedback.

4.4. Data Analysis

The accuracy of participants on this task was analyzed using a logistic regression model (Baayen 2008), with the *glm* package (Hothorn and Everitt 2014) in *R* (R

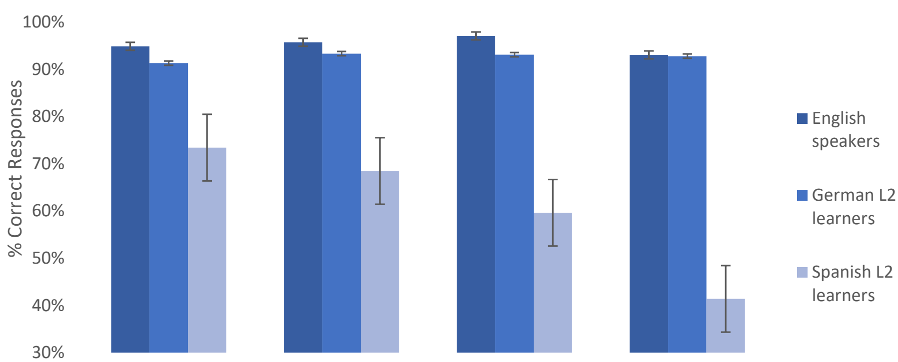
Development Core Team 2009). Three variables were considered as categorial predictors. The first one was L1, with three levels—English versus Spanish versus German—and with English serving as the baseline. Secondly, I measured the effects of the presence or absence of initial vowel—e.g., *estate* versus *state*—in the word to be monitored, and coded this variable as -0.5 if participants heard a vowel in the auditory stimulus and 0.5 if there was no vowel, as it was not clear which condition should be the baseline and which would potentially be more challenging for the participants. The third variable was the match/mismatch between the target word and the word in the auditory stimulus, with "Match" serving as the baseline.

As in Experiment 1, two sets of models were run—one on all accuracy rates using L1 as predictor and another on L2 learners' accuracy rates, with L1, proficiency and their interaction as predictors. The results of L2 proficiency are not reported because, after comparing models that included this variable with models that did not, the best fit was the model without proficiency. In each case, participant and item were included as random variables.

4.5. Results

Figure 2 presents the mean accuracy of the three L1 groups in the word-monitoring task and table 4 presents the results of the logistic regression model for all participants' accuracy.

FIGURE 2. Mean accuracy of the three L1 groups in the word monitoring task



Match condition		Mismatch condition	
Written: *state* Auditory: *state*	Written: *estate* Auditory: *estate*	Written: *state* Auditory: *estate*	Written: *estate* Auditory: *state*

TABLE 4. Logistic regression model on all participants' accuracy results

Variable	Estimate (SE)	z	p
(Intercept)	3.02 (0.2)	17.58	< 0.001
Match	0.03 (0.3)	0.14	> 0.1
Vowel	-0.19 (0.3)	< |1|	> 0.1
L1: German L2 learners	-0.03 (0.2)	-1.29	> 0.1
L1: Spanish L2 learners	-2.62 (0.2)	-14.01	< 0.001
Match x Vowel	1.1 (0.5)	2.19	< 0.05
Match x L1: German L2 learners	0.05 (0.3)	< |1|	> 0.1
Match x L1: Spanish L2 learners	0.06 (0.3)	< |1|	> 0.1
Vowel x L1: German L2 learners	-0.1 (0.4)	< |1|	> 0.1
Vowel x L1: Spanish L2 learners	-0.86 (0.4)	2.29	< 0.05
Match x Vowel x L1: German L2 learners	-0.77 (0.6)	–1.2	> 0.1
Match x Vowel x L1: Spanish L2 learners	-1.46 (0.5)	–2.68	< 0 .01

The results of the model summarized in table 4 show a main effect of L1 for the Spanish L2 learners of English, indicating a different pattern for this group compared with the English speakers. Secondary effects include the fact that the German learners did not differ from the native speakers; there was an interaction between vowel and L1 for Spanish learners, which shows that, unlike the native English speakers, they had different accuracies depending on whether words contained an initial vowel or not; and finally there was a three-way interaction between match, vowel and L1 for the Spanish speakers, indicating that their difficulty in identifying whether or not a vowel was present varied depending on whether the auditory word matched or mismatched the written word.

In order to better understand the pattern that emerged in the three-way interaction, two follow-up logit regression models were run to test the Spanish speakers' data for the effect of vowel separately for the match and mismatch conditions. These models are shown in tables 5 and 6 respectively.

TABLE 5. Logistic regression model on Spanish speakers' accuracy results, match condition

Variable	Estimate (SE)	z	p
(Intercept)	0.62 (0.06)	11.15	< 0.001
Vowel	0.08 (0.03)	2.41	< 0.02

TABLE 6. Logistic regression model on Spanish speakers' accuracy results, mismatch condition

Variable	Estimate (SE)	*z*	*p*
(Intercept)	0.6 (0.05)	12.85	< 0.001
Vowel	-0.15 (0.04)	4.12	< 0.001

The results show an effect of vowel in both conditions, match and mismatch. This means that in the match condition, the presence of a vowel hampers the identification of the corresponding word, that is, it results in a greater number of inaccurate responses (table 5). By contrast, in the mismatch condition, it is the absence of a vowel that yields higher levels of inaccuracy; in other words, it is when the competitor word with a vowel is activated that Spanish L2 learners have difficulty determining whether the target word contained a vowel (table 6).

4.6. Discussion

Experiment 2 was created with the aim of exploring whether, and if so, how, perception issues hamper word recognition. That is, a word-monitoring task was conducted to examine whether the perception difficulties reported in Experiment 1 had any effect on the learners' word recognition. There is a clear dichotomy in the results reported for Experiment 2. While the native English speakers and German L2 learners patterned similarly and showed no problems in detecting the target word in the auditory stimuli, the native Spanish speakers had difficulty detecting the target words, particularly in the mismatch conditions. Taken with precaution, these results seem to suggest that it is violations of abstract L1 syllable structure constraints rather than violations of L1 phonotactics that cause difficulties in L2 word recognition. Bearing in mind the differences between Spanish and German—Spanish does not allow any type of /s/-initial consonant cluster, while German allows sibilant-initial consonant clusters—if violations of L1 phonotactics were the cause of these misperceptions and word recognition difficulties, then both groups of learners would have patterned similarly. However, the results indicate that only the Spanish L2 learners experienced misperceptions and word recognition difficulties with respect to this specific parameter, which violates the syllable structure rule of their L1. These word recognition issues may be linked to the possibility that Spanish L2 learners of English are inappropriately activating lexical competitors when they hear words that contain an /s/-initial consonant cluster, as indicated by the results reported in the mismatch condition.

The possibility of L2 learners activating unintended lexical competitors has already been reported in the literature focusing on L1-L2 category assimilation and word recognition (Pallier et al. 2001; Weber and Cutler 2004; Escudero 2007;

Broersma and Cutler 2011). For the perception of spoken words, listeners need to match the incoming auditory stimuli with their stored lexical representations. In fact, word recognition models put forward the multiple parallel activation of lexical candidates, which then compete for selection (McClelland and Elman 1986; Marslen-Wilson 1987; Norris 1994). Hence, as the acoustic input unfolds, all the words that are consistent with it are active in the lexicon, until the individual receives enough segmental/suprasegmental information to successfully identify the intended word. In the perception of L2 words, not only words in the L2 but also in the L1 may be activated and compete for recognition. Thus, L1 phonotactics may result in the activation of competitor words that are not present in the signal, making L2 word recognition less efficient. One possible interpretation of the results of Experiment 2 is that the potentially perceived epenthetic vowel heard by the Spanish L2 learners of English may have activated the lexical competitor with an actual vowel. When the Spanish participants saw a word spelled with a vowel—e.g., *estate*—and then heard a mismatched sentence—e.g., *I have lived in this state for a long time*—their performance was below chance level as they correctly identified less than fifty percent of the trials. That is, when the Spanish speakers in this study heard a vowel-initial word, they activated both words—with or without a vowel—maybe because in their L2 lexicon they stored the /s/-initial words as either having or not having a vowel. This interpretation would therefore indicate that not only phonetic categories but also syllable structure constraints have an influence on L2 word recognition (Weber and Cutler 2004).

The vowel effect reported for the Spanish learners differed in the two match/mismatch conditions. While they were more accurate when listening to words that did not contain a vowel in the match condition, they were less accurate when they heard words that did not contain a vowel in the mismatch condition. Another possibility is that these participants were more sensitive to the written target, as they were more accurate whenever they saw a word without a vowel on the computer screen. While one possible interpretation of these results is related to the phonotactics of the participants' L1, it is also true that words without vowels were, in fact, more frequent in the experiment than words with a vowel due to the difficulty in finding (near-)minimal word pairs. That is, while /s/-initial words may be more difficult to perceive accurately, this disadvantage may be somewhat offset by their higher frequency in general conversations/texts, thus potentially influencing the results of the Spanish learners. The results of the second experiment may be driven by frequency instead of phonemic accuracy.

In summary, L2 syllable structure constraints—not only the exact sequences of segments allowed by L1 phonotactics—and the consequent lexical competition they induce seem to explain the pattern of results reported in Experiment 2, given that only the Spanish L2 learners of English showed these effects.

5. General Discussion and Conclusion

The current study tested proficiency-matched native Spanish and German L2 learners of English and a group of native speakers of English with the aim of determining whether L1-L2 syllable structure differences influence L2 spoken word recognition. Moreover, I explored how potential misperceptions might affect L2 word recognition by focusing on /s/-initial consonant clusters in L2 English.

In line with other studies that argue that L1 syllable structure constraints have an important effect on speech perception (Matthews and Brown 2004; Kabak and Idsardi 2007), the results of Experiment 1—the AXB task—show that only Spanish L2 learners of English had difficulties in discriminating between /s/-initial clusters with or without an epenthetic vowel. Importantly, these syllable structure constraints seem to be abstract in nature, as their influence on L2 speech perception is not limited to the precise sequencing of consonants in the L1 and the L2, as is evidenced by the pattern of results for the German group. As described earlier, German does not allow /s/-initial consonant clusters but it does allow sibilant-initial clusters. Thus, it is the sequencing of sounds that prevents the perception of epenthetic vowels in German L2 learners of English, but not in the Spanish learners, who seem to rely on vowel epenthesis to repair this violation of L1 phonotactics. This pattern is consistent with how the two languages adapt loanwords into their lexicons: German incorporates /s/-initial consonant clusters by turning the alveolar fricative into a palatal fricative—e.g., *stoppen* [ˈʃtɔ.pən] "to stop" (Itô and Mester 2001)—whereas Spanish requires the insertion of a vowel before the consonant cluster—e.g., *eslogan* [esˈlo.gan] "slogan" (Diccionario de la Real Academia Española 2009).

Research focused on understanding the effect of L1 syllable constraints on L2 word recognition is still limited in nature in comparison with the extensive research on how L1-L2 differences between phonemic repertoires can lead to speech perception problems. Experiment 2 was specifically designed to investigate how L2 learners' perceptual difficulties affected their L2 word recognition by using a word-monitoring task. This task was chosen because it is considered to be a direct measure of participants' unconscious access to their mental representations (Kilborn and Moss 1996). The results of the current study reinforce the idea that it is not only differences in L1-L2 phonetic categories but also L1 syllable structure constraints that influence L2 word recognition. In line with previous evidence, this suggests that learners may activate unintended—or phantom—lexical competitors, which may lead to perceptual and, thus, lexical confusion (Weber and Cutler 2004; Broersma and Cutler 2011). While this may be the case, it is also true that in a real-word situation, listeners have access not only to bottom-up acoustic phonetic information, but also to top-down cues—grammatical or contextual. In the current study, the sentences were created such that they were semantically ambiguous so that only bottom-up acoustic phonetic information could help disambiguate between the two forms tested—with versus without a vowel. However, future studies should examine whether other types of cue—for example, semantic or syntactic—interact with bottom-up acoustic phonetic information in constraining lexical activation in L2 speech processing.

Another point that should be taken into account when interpreting the results of both experiments in this study is that they use very different sets of stimuli. While Experiment 1 included nonce words, Experiment 2 used real minimal pairs. As discussed for Experiment 2, some of the lexical effects reported could be explained by reference to differences in the overall frequency of the individual words in the minimal pair. Indeed, certain interesting effects have been interpreted as evidence for the top-down modulation of lower-level auditory perception (Myers and Blumstein 2008; Shahin et al. 2009). In the current study the pattern of results of the two tasks are in line with each other, indicating that the same process may be at play in both. However, future research should address this discrepancy in the type of stimuli used in the study.

Finally, while in the present study proficiency was not found to statistically explain the results obtained, this should be taken with caution. Although the findings reported here may suggest that the effects of L1 syllable structure on L2 speech perception exist across proficiency levels and persist even at advanced levels, other interpretations are also possible. The proficiency cloze test used in this study to determine the proficiency of the participants may not have been an adequate measure of aural capability. As this cloze test was conducted in the visual modality, it may not have been sensitive enough to capture variability in L2 learners' perceptual skills, and thus it may have masked any potential proficiency effects. Future research should examine the extent to which native-like production, perception and recognition of /s/-initial consonant clusters is acquirable for Spanish L2 learners whose proficiency is assessed aurally.

Works Cited

Abrahamsson, Niclas. 1999. "Vowel Epenthesis of/sC (C)/Onsets in Spanish/Swedish Interphonology: A Longitudinal Case Study." *Language Learning* 49 (3): 473-508.

Anderson, Janet. 1987. "The Markedness Differential Hypothesis and Syllable Structure Difficulty." In Ioup and Weinberger 1987, 279-91.

Baayen, Harald R. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge UP.

Best, Catherine T. 1995. "A Direct Realist View of Cross-Language Speech Perception." In Strange 1995, 171-206.

Best, Catherine T., Gerald W. McRoberts and Nomathemba M. Sithole. 1988. "Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-speaking Adults and Infants." *Journal of Experimental Psychology: Human Perception and Performance* 14 (3): 345-60.

Best, Catherine T. and Michael D. Tyler. 2007. "Nonnative and Second-Language Speech Perception: Commonalities and Complementarities." In Bohn and Munro 2007, 13-34.

BOHN, Ocke-Schwen and Murray J. Munro, eds. 2007. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Amsterdam and Philadelphia: John Benjamins.

BROERSMA, Mirjam and Anne Cutler. 2011. "Competition Dynamics of Second-Language Listening." *Quarterly Journal of Experimental Psychology* 64 (1): 74-95.

BROSELOW, Ellen and Daniel Finer. 1991. "Parameter Setting in Second Language Phonology and Syntax." *Interlanguage Studies Bulletin* 7 (1): 35-59.

BROWN, James Dean. 1980. "Relative Merits of Four Methods for Scoring Cloze Tests." *The Modern Language Journal* 64 (3): 311-17.

CARLISLE, Robert S. 1997. "The Modification of Onsets in a Markedness Relationship: Testing the Interlanguage Structural Conformity Hypothesis." *Language Learning* 47 (2): 327-61.

—. 1998. "The Acquisition of Onsets in a Markedness Relationship: A Longitudinal Study." *Studies in Second Language Acquisition* 20 (2): 245-60.

DAVIDSON, Lisa. 2006. "Phonology, Phonetics or Frequency: Influences on the Production of Non-Native Sequences." *Journal of Phonetics* 34 (1): 104-37.

DAVIDSON, Lisa, Peter Jusczyk and Paul Smolensky. 2004. "The Initial and Final States: Theoretical Implications and Experimental Explorations of Richness of the Base." In Kager, Pater and Zonneveld 2004, 321-68.

DEHAENE-LAMBERTZ, Ghislaine, Emmanuel Dupoux and Ariel Gout. 2000. "Electrophysiological Correlates of Phonological Processing: A Cross-Linguistic Study." *Journal of Cognitive Neuroscience* 12 (4): 635-47.

DICCIONARIO DE LA REAL ACADEMIA DE LA LENGUA ESPAÑOLA (RAE). 2009. s.v. "Eslogan." [Accessed February 10, 2021].

DUPOUX, Emmanuel et al. 1999. "Epenthetic Vowels in Japanese: A Perceptual Illusion?" *Journal of Experimental Psychology: Human Perception and Performance* 25 (6): 1568-78.

— et al. 2001. "New Evidence for Prelexical Phonological Processing in Word Recognition." *Language and Cognitive Processes* 16 (6): 491-505.

ECKMAN, Fred R. and Gregory K. Iverson. 1993. "Sonority and Markedness among Onset Clusters in the Interlanguage of ESL Learners." *Second Language Research* 9 (3): 234-52.

ESCUDERO, Paola. 2007 "Multilingual Sound Perception and Word Recognition." *Stem- Spraak-en Taalpathologie* 15 (2): 93-103.

FLEGE, James Emil. 1995. "Second Language Speech Learning: Theory, Findings and Problems." In Strange 1995, 233-77.

FLEGE, James Emil, Murray J. Munro and Ian R. A. MacKay. 1995. "Factors Affecting Strength of Perceived Foreign Accent in a Second Language." *The Journal of the Acoustical Society of America* 97 (5): 3125-34.

FLEGE, James Emil, Grace H. Yeni-Komshian and Serena Liu. 1999. "Age Constraints on Second-Language Acquisition." *Journal of Memory and Language* 41 (1): 78-104.

Hancin-Bhatt, Barbara and Rakesh M. Bhatt. 1997. "Optimal L2 Syllables: Interactions of Transfer and Developmental Effects." *Studies in Second Language Acquisition* 19 (3): 331-78.

Hothorn, Torsten and Brian S. Everitt. 2014. *A Handbook of Statistical Analyses Using R*. 3rd ed. Boca Ratón, FL: Chapman and Hall/CRC Press.

Ioup, Georgette and Stephen Weinberger, eds. 1987. *Interlanguage Phonology: The Acquisition of a Second Language Sound System*. Cambridge, MA: Newbury House.

Itô, Junko and Armin Mester. 2001. "Covert Generalizations in Optimality Theory: The Role of Stratal Faithfulness Constraints." *Studies in Phonetics*, *Phonology and Morphology* 7 (2): 273-99.

Kabak, Bariş and William J. Idsardi. 2007. "Perceptual Distortions in the Adaptation of English Consonant Clusters: Syllable Structure or Consonantal Contact Constraints?" *Language and Speech* 50 (1): 23-52.

Kager, René, Joe Pater and Wim Zonneveld, eds. 2004. *Constraints in Phonological Acquisition*. Cambridge: Cambridge UP.

Kilborn, Kerry and Helen Moss. 1996. "Word Monitoring." *Language and Cognitive Processes* 11 (6): 689-94.

Marslen-Wilson, William D. 1987. "Functional Parallelism in Spoken Word-Recognition." *Cognition* 25 (2): 71-102.

Martinez-Garcia, Maria Teresa. 2018. "The Effect of 'Illusory Vowels' in Spanish-Speaking Second Language Learners of English." *Language and Linguistics* 79: 147-76.

Martinez-Garcia, Maria Teresa and Annie Tremblay. 2015. "Syllable Structure Affects Second-Language Spoken Word Recognition and Production." In Wolters et al. 2015, n.p.

Matthews, John and Cynthia Brown. 2004. "When Intake Exceeds Input: Language Specific Perceptual Illusions Induced by L1 Prosodic Constraints." *International Journal of Bilingualism* 8 (1): 5-27.

McClelland, James L. and Jeffrey L. Elman. 1986. "The TRACE Model of Speech Perception." *Cognitive Psychology* 18 (1): 1-86.

Myers, Emily B. and Sheila Blumstein. 2018. "The Neural Bases of the Lexical Effect: An fMRI Investigation." *Cerebral Cortex* 18 (2): 278-88.

Norris, Dennis. 1994. "Shortlist: A Connectionist Model of Continuous Speech Recognition." *Cognition* 52 (3): 189-234.

Pallier, Christophe, Angels Colomé and Núria Sebastián-Gallés. 2001. "The Influence of Native-Language Phonology on Lexical Access: Exemplar-Based versus Abstract Lexical Entries." *Psychological Science* 12 (6): 445-49.

R Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Shahin, Antoine, Christopher Bishop and Lee Miller. 2009. "Neural Mechanisms for Illusory Filling-in of Degraded Speech." *NeuroImage* 44 (3): 1133-43.

STRANGE, Winifred, ed. 1995. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore, MD: York Press.

TAGLIAFERRI, Bruno. 2005. *Paradigm*. Perception Research Systems, Inc.

WEBER, Andrea and Anne Cutler. 2004. "Lexical Competition in Non-Native Spoken-Word Recognition." *Journal of Memory and Language* 50 (1): 1-25.

WOLTERS, Maria et al. eds. 2015. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: The Scottish Consortium for ICPhS 2015.

Received 14 February 2020 Revised version accepted 9 March 2021

Maria Teresa Martinez-Garcia completed her PhD in linguistics at the University of Kansas in 2016. Her research interests focus on experimental linguistics, particularly bilingualism and second language speech perception and production. Currently she is as an assistant professor in the World Languages & Cultures Department at the University of Utah, Asia Campus.