Resolving contiguous vowels across word boundaries in Spanish: L2 learners, levels, and tasks*

Erik W. Willis (Bloomington), D. Eric Holt (Columbia) and Carly Carver (Augusta)

Abstract

This paper reports on the application of synalepha, the phonological process of merging or linking contiguous vowels across word boundaries, as in *mi amigo* [mi̯a.ˈmi.ɣ̟o], meaning 'my friend'. We examine two English-speaking L1 groups learning Spanish as an L2 compared to a Spanish monolingual-speaking group and include a narrative-retell and a reading task. In the reading task, the intermediate learner group produced linked or connected vowel sequences in only 30% of the potential contiguous sequences, compared to a frequency of linked vowels at 74% by advanced graduate student learners. The intermediate learners increased the frequency of linking contiguous vowels to 49% in the narrative-retell task, whereas the advanced L2 speakers slightly reduced their linking rate to 68%. The Mexican monolingual speaker comparison group produced linked speech in a narrative-retell task at a rate of 93%. These findings show that the application of phonological processes is impacted by task type and has implications for classroom teaching activities that incorporate narrative-retell tasks to encourage monolingual patterns of vowel linking.

1 Introduction

In Spanish, vowels across word boundaries (V#V) are characterized as being resolved by monolingual speakers in a smooth manner with full linking without a separation between the two vowels (Whitley 2002). In English, however, glottal stops tend to intervene, which textbooks warn not to transfer and instead to produce "smooth linking", the putative monolingual norm (cf., Bowen/Stockwell 1960; Stockwell/Bowen/Martin 1965; Whitley 2002; Morris 2010). Additional types of linking have been observed and described (cf. Alba 2006; Face/Alvord 2004; González/Weissglass 2015; Hualde/Prieto 2002; Smith/Flores/Gradoville 2008) and include diphthong formation, merged/hybrid vowels, replaced/eliminated vowels, and degrees of shortening when V1=V2. Despite the descriptions, they have received little empirical attention or documentation (cf. Alba 2006 2008; Face/Alvord 2004; Hualde/Prieto 2002; Smith/Flores/Gradoville 2008). For Spanish as a Second Language (L2 Spanish), while some of these types

* We wish to acknowledge and express our appreciation to the anonymous reviewers, the editors, and the attendees of L2APR: L2 Accent and Pronunciation Research: Acquisition, Teaching, and Attitudes at Ca' Foscari University of Venice, as well as Macy Maas and Magaly Gastelum from Indiana University, for their valuable comments and suggestions. The usual disclaimers apply to the content.

are treated in pedagogical texts, strategies such as merged and replaced vowels are rarely mentioned.

The current study investigates V#V production of Spanish monolingual speakers and second language learners of Spanish, examining the distribution of continuous or linked speech and exploring various resolution types (see Figures 2 and 3 for examples) based on vowel formant structure, vowel duration(s) and auditory impression. We also explore the extralinguistic factors of L2 level and task type.[1]

When vowels come into contact in Spanish across a word boundary (V#V), they may connect or link smoothly in various ways in monolingual speech, including a diphthong, diphthong from a mid vowel glide, a merged same vowel,[2] a replaced vowel, and hiatus:[3]

(1)

Synalepha (pronounced in the same syllable)		
diphthong formation	*la historia* 'the history'	[laɪ̯s.ˈto.ɾɪ̯a]
gliding of mid vowel	*le habla* 'speaks to him	[lɛ̯a.ˈβla], [liɪ̯a.βla][4]
merged same vowels	*lo honró* 'he/she honored'	[lon. ˈro]
replaced vowel	lo importante	[lim.poɾ.'taɲ̟.te]
Hiatus (pronounced in separate syllables)		
hiatus	*canta ella* 'she sings'	[ˈkaɲ̟.ta.ˈe.ja]

The connected, linked, or continuous speech resolution of contiguous vowels contrasts with separated or discontinuous sequences in which there is creaky voice, a glottal stop, or a pause between the vowel sequence.

2 Literature review

2.1 Synalepha in Monolingual Spanish

It has been claimed that synalepha occurs categorically in monolingual Spanish speech (cf. Colina 2009; Harris 1983; Real Academia Española 2011; Stockwell/Bowen/Martin 1965; Whitley 2002) as it is the preferred way to pronounce two adjacent vowels across word boundaries. Additionally, there have been rules proposed by certain authors that contradict each other regarding how vowel sequences between words are resolved. For example, according to Martínez-Gil (2000: 517), in Chicano Spanish, a mid vowel (whether stressed or unstressed) is elided when it precedes a high vowel. In contrast, Jenkins' (1999: 52f.) claims that stressed vowels are never modified unless they are adjacent to another stressed vowel in New Mexican Spanish, and hiatus is maintained when a stressed vowel precedes a vowel other than /e/. Therefore, considering the examples shown in (1), the resolution of the hiatus will be different depending on the rules proposed by each dialect. For the data treated by Martínez-Gil (2000), the

[1] In future work, we address linguistic factors such as vowel quality, stress, and word type.

[2] There were a few cases of another merged vowel type, typically produced as a centralized vowel or a vowel glide between two vowels, which was very short; for example, *la elefante* 'the elephant' → [lɛ.le.'faɲ̟.te].

[3] This merger often results in the merger of the two syllables into a single syllable (a form of resyllabification). Importantly, linked speech as hiatus does not impact the syllable structure. Consonantal resyllabication in which a word final consonant resyllabifies with a vowel in the following word is another manifestion of word linking.

[4] Mid vowels will often rise to a high vowel creating a synalepha (cf. Quilis 1993: 190f. for a full discussion).

/e/ would be elided, while for that of Jenkins (1999), the /e/ would be maintained. However, Figure 1 is an example of a two-word sequence that illustrates a diphthong not predicted by rule-based claims.

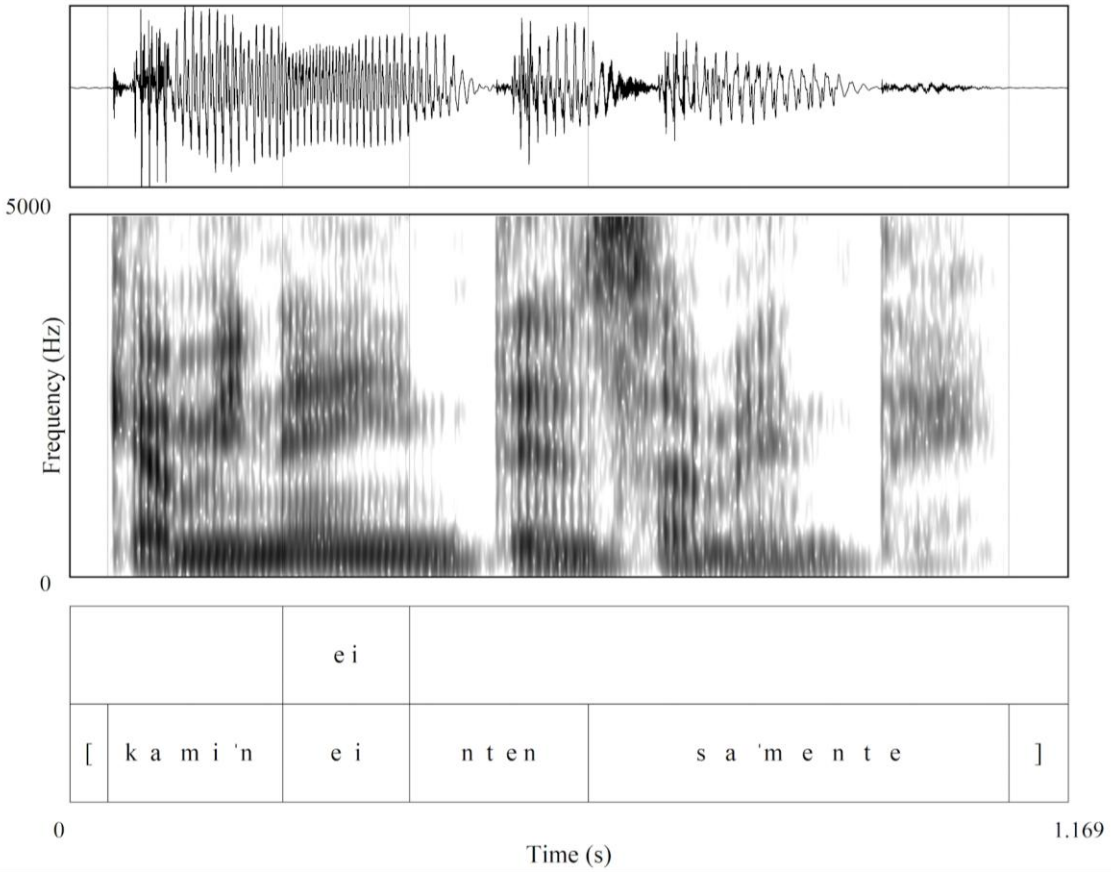

Figure 1: Sample spectrogram of the words *caminé intensamente* 'I walked intensely' by a native Spanish speaker from southern Spain

There are distinct goals in examining the apparent contrast between strict rule-based explanations of hiatus resolution, which seek to uncover underlying representations (cf. Hualde/Prieto 2002; Chitoran/Hualde 2007), and the descriptive goals that emphasize social variables or the role of frequency in sound change (cf. Alba 2006). While our goal is likewise descriptive, our focus also diverges slightly from Alba (2006), who investigates synalepha as a reduction process within a frequency-based model of lenition. Instead, we emphasize the linking or connectedness of contiguous vowel sequences as a phonological process of Spanish with multiple types of connected speech manifestations that English-speaking L2 learners of Spanish produce variably. We consider both hiatus and synalepha as forms of connected speech that function within monolingual patterns.

While some studies have explored this process of hiatus resolution empirically and investigated factors such as frequency effects, many rely on authors' and participants' intuitions or impressionistic analyses (cf. Hualde/Prieto 2002; Chitoran/Hualde 2007; Alba 2006, 2008). These approaches often lack the support of acoustic or spectral analysis, which is critical for determining the accurate patterns of synalepha among monolingual Spanish speakers. In the following section, studies investigating hiatus resolution among monolingual Spanish speakers, including heritage speakers from various dialects, will be discussed.

The most thorough examination of contiguous vowel production across word boundaries is Alba (2005, 2006), who examined hiatus resolution or maintenance in New Mexican and Colorado Spanish. The data consisted of 29 hours of sociolinguistic interviews and a lexical identification task with 38 native speakers, both men and women. However, the data from the lexical identification task were not used because the author wanted to report results for data representative of natural discourse. Alba defined his variable context of synalepha as between the vowel /a/ and any vowel at the word boundary, excluding several contexts, including if "there was any kind of pause or stuttering between the two words in question" (Alba 2006 276). Impressionistic coding was used to determine whether a case represented hiatus maintenance or resolution due to the author's assertion that spectral analysis leads to difficulties in measuring vowel quality and boundaries as formant frequencies (F1 and F2) present much variation. Additionally, although a variety of resolution strategies are employed among natives (previously enumerated in (1)), Alba only coded for hiatus resolution or maintenance since impressionistic analysis does not lend itself to a clear identification of these resolution strategies. The following factors were considered in his analysis: the stress of each vowel, the type of syllable (open or closed) in which the second vowel in the sequence occurred, the word class (content or function), prior mention of the two words, and the relative frequency of the words in question.

Out of 1,912 occurrences considered in the corpus, the results revealed a higher percentage of hiatus resolution than maintenance (74% vs. 26%), with more resolution (83%) when both vowels are unstressed and 13% in stressed syllables. Alba (2006) did not investigate the potential effect the task might have on hiatus resolution since he did not include the lexical identification task data in his analysis. However, differences according to task may advance understanding of hiatus resolution strategies and are investigated in this study.

Monolingual Spanish was explored by Smith/Flores/Gradoville (2008), who analyzed vowels in hiatus context at word boundaries in Veracruz, Mexico Spanish, to investigate the specific conditions under which hiatus maintenance and different resolution types (diphthong, vowel reduction or elision, or creation of a new vowel) occur. Specifically, they investigated whether transitions between the two vowels follow a consistent pattern or if there are variable articulations. They also examined the duration of vowel sequences. The data is from a corpus by Willis (2008) of semi-spontaneous narrated speech. The informants provided a narrative telling of the wordless picture book *Frog, Where Are You?* by Mercer Mayer (1969). The speakers were aged 18–28 and were all university or professional students. Acoustic and spectral analysis using Praat (cf. Boersma/Weenink 2021) was employed to plot the vowels, measure the duration of the vowel sequence, and describe the resolution. The longer vowel trajectories and greater duration indicated less resolution. The results revealed 94% hiatus resolution with various resolution types, such as vowel fusion, replacement of one vowel with another, and even centralization among some speakers. Regarding duration, less resolution was associated with longer sequence duration. Regarding other factors contributing to hiatus resolution or maintenance, it was found that syllabic stress did not impact hiatus resolution in contrast to the findings of Alba (2006).

Speech rate and word frequency also was found to play a role in the rates of hiatus resolution in Puebla Mexico Spanish (cf. Souza 2010), and he argues for a usage-based model of phonology similar to Alba (2006). Barbería Aurrekoetxea (2012) reports that informants from the

northern part of Spain tended to maintain hiatus forms within words and tended to diphthongize across word boundaries in read speech.

When synthesizing the previously discussed studies, some common findings and patterns are noted. First, the use of acoustic and spectral analysis is found in relatively few investigations of hiatus resolution, resyllabification, and synalepha, where impressionistic analysis is normally employed. Second, although it has been consistently found that there is more hiatus resolution, resyllabification, and synalepha among monolinguals than preferences for other strategies, some variation has still been found among monolinguals. Still, this variation in resolutions is not usually explained descriptively in terms of specific strategies nor empirically using more advanced statistical models to identify factors that favor and disfavor these processes among monolinguals.

2.2 Synalepha and similar processes: previous literature on L2 Spanish learners

Among studies describing hiatus maintenance and resolution patterns through the use of synalepha and the factors conditioning these phenomena among monolingual Spanish speakers, there has been little research on the patterns of these phenomena in the speech of Spanish second language learners (L2). American English is reported to have continuous transitions between vowels across word boundaries regularly (cf. Olive/Greenwood/Coleman 1993). However, they also state that "It is not uncommon for speakers to insert a glottal stop between the articulation of the two vowels to signal a word boundary" (ibid: 120). Later, they state that "Discontinuous transitions are caused by the presence of a glottal stop" (ibid: 131). The pronunciation literature for English-speaking learners of Spanish is more assertive regarding the carryover of separated vowel sequences across word boundaries. In textbooks for Spanish L2 learners who are native English speakers, the use of synalepha is prescriptively encouraged to Spanish L2 learners instead of transferring English strategies like glottal stops (cf. Davidson/Erker 2014; Whitley 2002) and consonantal insertion (cf. Bowen/Stockwell 1960); these strategies occur in English with adjacent vowels at the word boundary, "one vowel never follows another [...] without a consonant, a semivowel, or, under special conditions, a glottal stop between the two vowels" (Bowen/Stockwell 1960: 108). According to Whitley (2002: 32): "even when two vowels meet in the context of a hiatus, they should be smoothly linked in Spanish [...] most varieties of Spanish require complete linking between all consonants and vowels in a phrase".

In a related phenomenon, Holt/Gelormini-Lezama (2008) analyzed resyllabification and synalepha among native English speakers learning Spanish as L2. Resyllabification is defined as the phenomenon where there is a link between a word that ends in a consonant and another that begins with a vowel, changing the syllabic structure of these segments to yield the preferred syllabic structure in Spanish (CV) as in *los ojos azules* [lo.ˈso.xo.sa.ˈsu.les] (cf. Beristain 2024; Jiménez-Bravo/Lahoz-Bengoechea 2023; Marchini/Ramsammy 2023; Holt/Gelormini-Lezama 2008). Since resyllabification and synalepha are optional in English, their acquisition is assumed to be difficult for learners of Spanish. Three treatments were carried out: study abroad, study with explicit instruction, and study in another advanced Spanish course in business or literary analysis in which there was no explicit instruction or practice of pronunciation. There were four groups in the study. Group A had 19 participants who took an introductory course on

phonetics and pronunciation. Group B consisted of 15 participants who participated in a four-week study program in Costa Rica. Group C had 22 participants who took another advanced class without explicit instruction on the two phenomena. Group D was a group of 6 native speakers from Costa Rica, Mexico, Equatorial Guinea, Venezuela, and Puerto Rico. Data were elicited using a list of 100 sentences containing vowel combinations to examine. All learners read one version of the list at the beginning (time 1) and the end (time 2) of the treatment, and the order was counterbalanced. The monolinguals only did the task once.

One of the main objectives of their study was to examine how different modes and types of exposure to Spanish and the effect of explicit instruction on pronunciation accuracy and to document the prevalence of connected speech among native speakers. It was predicted that due to the focused practice on pronunciation and greater exposure to monolingual speech with a greater need for oral expression, Group A and Group B would show the greatest improvement, respectively. Impressionistic coding was used to determine cases of synalepha and resyllabification. The monolingual speakers produced synalepha in 74.52% of the cases.

It is evident that the percentage of synalepha among learners is less than among monolinguals, that explicit pronunciation instruction led to the greatest improvement in synalepha among learners, and that native monolingual speakers did not produce synalepha categorically, as it only occurred 74.52% of the time, one of the lowest rates seen in the examination of previous studies. However, no statistical tests were implemented to determine if the results were significant or random, no details were provided regarding the other strategies participants used instead of resyllabification or synalepha, and other factors that could impact synalepha and what favors or disfavors it were not examined. Finally, no acoustic and spectral evidence was used to identify the cases of synalepha. The authors acknowledge these limitations and offer several future directions for this topic that lie beyond the scope of the current paper. The current study is similarly exploratory, with the inclusion of exploratory but includes acoustic details to develop a deeper understanding of the processes involved.

González/Weissglass (2015) investigated the effect of task type and instruction on resyllabification with the goal of also noting pedagogical implications based on data from only two utterances. The participants were 25 Spanish learners whose native language is English (21 women and four men aged between 18–22 years) with an intermediate to advanced proficiency level and eight native speakers (five women and three men aged between 20–24 years) from various dialects including Peruvian, Spanish, Cuban, Colombian, Venezuelan, and Bolivian, who had been in the United States for 0–14 years. The learners were divided into two groups: one that received 10 minutes of explicit pronunciation instruction at the end of each day of their regular class and another group that did not receive such instruction. The instruction consisted of the basics of articulatory phonetics, visual and aural demonstration of differences between sounds, and guided practice. Two tasks were used in this study: a sentence reading task and a picture description task. The learners completed both tasks at the beginning and end of the semester, while the native speakers completed them once. Praat was used to analyze the speech where discontinuous transitions indicated a lack of resyllabification and continuous transitions indi-

cated resyllabification. The duration of the transition where resyllabification should occur relative to the total duration of the two segments was also measured.[5]

It was predicted that learners would produce less resyllabification than natives since this phenomenon occurs very rarely and is optional in English. Additionally, it was predicted that there would be more discontinuous transitions (such as glottal stops and creaky voice) in the speech of learners than among natives since glottalization is a very common articulatory process in native English (cf. González/Weissglass 2015). Finally, regarding duration, it was predicted that the duration of the transition would be shorter where there is resyllabification and longer where there is no resyllabification. A total of 1,392 occurrences were analyzed. They found less resyllabification among learners than natives (33% vs. 72%, respectively). Regarding the strategies used by learners to produce contiguous vowels across word boundaries, there were found many instances of glottal stops and creaky voice. The average duration of resyllabification among learners was 122 ms, and that of glottal stops was 411 ms. Regarding differences according to task type, more resyllabification was found in the picture description task than in the reading task (30% vs. 20% respectively), which the authors attribute to a potential orthographic effect in the sentence reading task, i. e., the physical space between written words led to a higher rate of discontinuous transitions. They also mention the inclusion of the letter <h> and that all speaker groups, including natives, produced less resyllabification in this context. These results also relate to the fact that most pronunciation instruction is based on reading. Regarding the impact of instruction in this study, there was no improvement in the picture description task; however, in the reading task, it was found that explicit pronunciation instruction increased resyllabification by 12%, while the group that did not receive instruction decreased their resyllabification performance by 7%. Overall, the authors conclude that there is very little resyllabification among learners and that there is a need for earlier pronunciation instruction in Spanish classes and a greater variety of tasks such as word and sentence reading, picture descriptions, etc.

In sum, few studies have detailed the strategies used to resolve hiatus or produce synalepha across word boundaries (cf. Smith/Flores/Gradoville (2008) and Barbería Aurrekoetxea (2012) for exceptional cases). There are limited studies characterizing synalepha patterns among monolingual speakers and Spanish learners, and even fewer that compare not only frequency patterns between the two groups and comparisons of the different factors conditioning synalepha patterns in spontaneous or continuous speech. Finally, there has been no comparison among learner levels beyond intermediate, and the role of task effects has not been explored. Additional processes of sound patterns across word boundaries have shown variability in the acquisition of the processes reported as productive in L1 Spanish.

It is important to examine potential variation in monolingual speech of any structure when addressing the acquisition and teaching of these structures. The acquisition of variable structures has been documented in Spanish for morphosyntactic structures (e. g., mood and copula contrast and subject expression (cf. Geeslin 2011). However, the acquisition of phonological and phonetic variation has received less attention, especially those examining suprasegmental

[5] We note that this definition of resyllabification may be problematic, and we prefer to limit our characterization to linked speech versus discontinuous or separated sequences.

aspects like synalepha (cf. Holt/Gelormini-Lezama 2008), and to our knowledge, no study has examined the variable nature of synalepha or linked speech among Spanish L2 learners from a variationist perspective. These studies on the acquisition of variation in L2 Spanish highlight the importance of thoroughly understanding the monolingual system when establishing norms for comparison with the acquisition and use of the target language by learners (cf. Geeslin 2003). This idea is revisited in the discussion section of the results.

2.3 Research questions

The previous review of the literature leads us to the following research questions:

1. How do English L1-speaking learners of Spanish resolve contiguous vowels across word boundaries compared to monolingual speakers?
2. Are there patterns of resolution strategies among the speaker groups (monolinguals, advanced learners, intermediate learners)
   - Continuous speech/smooth linking (synalepha and hiatus)?
   - Discontinuous speech/separation (glottal stops, creak)?
3. Do resolution strategies of contiguous vowels across word boundaries vary by task type?

3 Methodology

3.1 Participants

The current study examines the production of contiguous vowels across word boundaries among several groups of speakers and two task types. Two learner groups, intermediate (fourth-semester language students) and advanced (graduate students), are studied and compared to documented monolingual norms from Veracruz, Mexico.[6] Learner data come from a simple picture narrative (40–50 token sequences) and a short reading passage (59 vowel-to-vowel contexts).

The monolingual Spanish speakers were four females from Veracruz, Mexico, pursuing professional or university studies, aged 18–28, and were recorded in June 2005 in a quiet room in Veracruz, Mexico, by the first author.[7] All learners were native English speakers and included an equal number of males and females for the advanced learner graduate students. The eight intermediate students were primarily female and were all Spanish majors or minors aged between 19 and 23 at a large midwestern university in the United States. The advanced learners were part of the same department, and their ages ranged from 23 to 32. The third author recorded them in a linguistics studio with a sound-attenuated room in 2014–2015.

All speakers were recorded with a USBpre sound device sampled at 44K and a Shure WH10A microphone with the condenser microphone between 1/2 and 3/4 inches from their lips. The sound was then examined using the acoustic software Praat (cf. Boersma/Weenink 2021).

---

[6] Indiana University Institutional Review Board STUDY #1510443387

[7] The data collection with the Veracruz Mexican speakers was conducted prior to the conceptualization of the current project and we only have narrative data for comparison.

Speaker data is presented in Table 1.

Population	Speaker-group	Number of speakers	Tokens	Task
Monolinguals	Veracruz, México	4 females	279	Frog picture narrative
Advanced (Graduate students)	Midwest US	8 (4 males x 4 females)	828	Frog picture narrative/ GM-para reading
Intermediate (Undergraduates)	Midwest US	8 (7 females x 1 male)	842	Frog picture narrative/ GM-para reading

Table 1: Informant groups and tasks (GM=García-Márquez paragraph reading text)

3.2 Tasks

The data collection for the study involved three tasks across different informant groups. It began with a short sociolinguistic interview, followed by a narrative retelling of Mercer Mayer's wordless picture book, *Frog, Where Are You?* For monolingual speakers, the contiguous vowels or synalepha data was based on this narrative retelling task. Learner participants also started with a short sociolinguistic interview, which included background characteristics, and then completed the narrative retelling of the Frog story. Finally, they completed a reading task consisting of an excerpt about Gabriel García Márquez from a second-year basic language course.

This reading selection was chosen because all intermediate learners were expected to have been exposed to the textbook and its materials in previous coursework, ensuring the vocabulary was within their exposure level. Advanced learner participants, who may have taught the course, were assumed to find the reading level appropriate given their familiarity and multiple exposures to the material.

3.3 Analysis

All three original investigators were involved in the data analysis and followed a particular coding schema.[8] The first level of analysis was whether or not the vowel sequence was linked and was characterized in a binary distinction as a linked or separated sequence.

The analysis of the linked status, used as a proxy for synalepha writ large, was based on acoustic analysis and characterization based on multiple acoustic cues. In all cases, we first identified the onset of the first vowel to the end of the second. The next level of analysis was the type of linking or how the vowels were separated, based on several acoustic cues that included the sequence duration and formant values and patterns described in section 4.

[8] We included details in the data collection on the preceding and following vowels, stress, and lexical class that are beyond the scope of the current paper.

4 Results

4.1 General patterns

In the examination of vowel sequences across word boundaries, there were clear resolution patterns for linked vowels versus separated vowels. The distinction between a connected and a separated vowel sequence was the easiest to determine. No single cue was used invariably. Figure 2. exemplifies a token of linked vowels across word boundaries in the phrase *y buscaba y gritaba*, 'and he was searching and shouting', with no breaks or discontinuities between them and a duration of 93 ms.

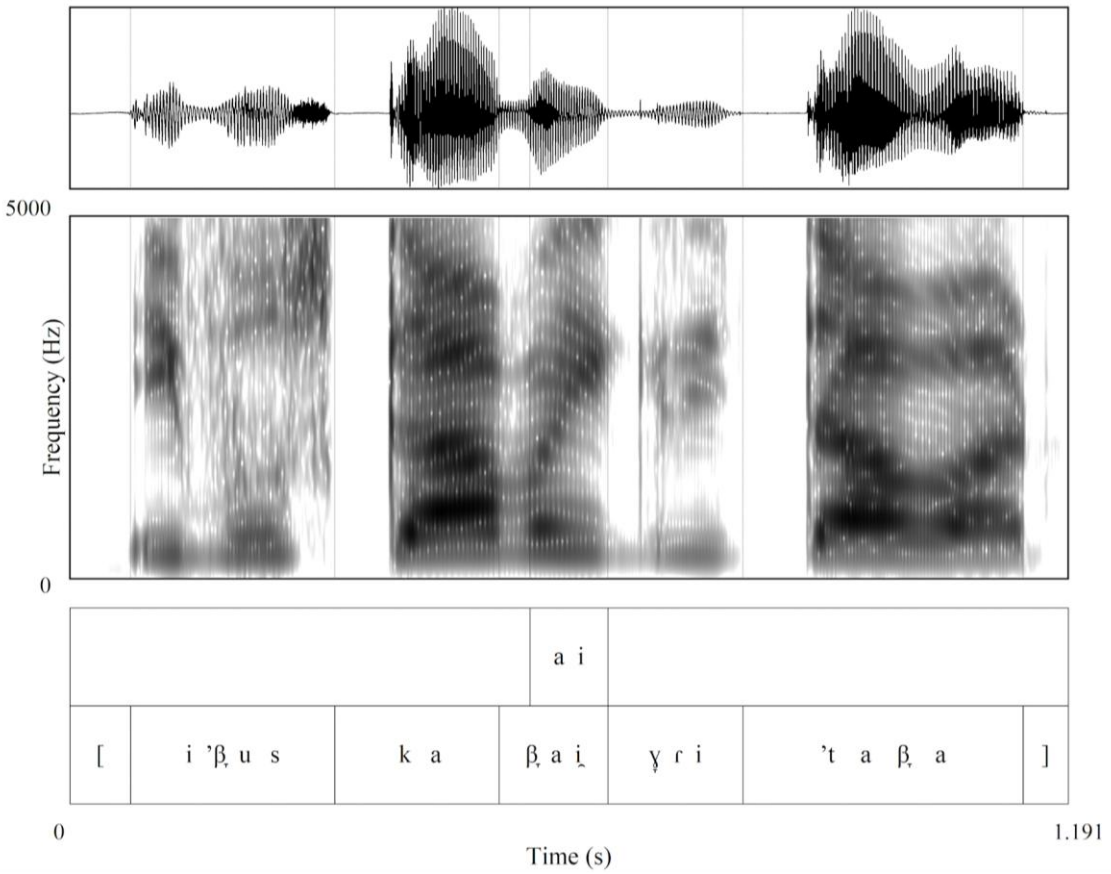

Figure 2: Linked vowels (93 ms) in the word sequence *y buscaba y gritaba*
'he was searching and shouting'

The category of resolution described as "linked" or "connected" speech had no visible or appreciable break in the waveform or in the spectrogram. A vowel sequence separation was determined by the presence of a creak or glottal fry, a glottal stop, or a pause between the two vowels.

The most commonly occurring separation strategy for learners was the insertion of a glottal stop, exemplified in Figure 3.

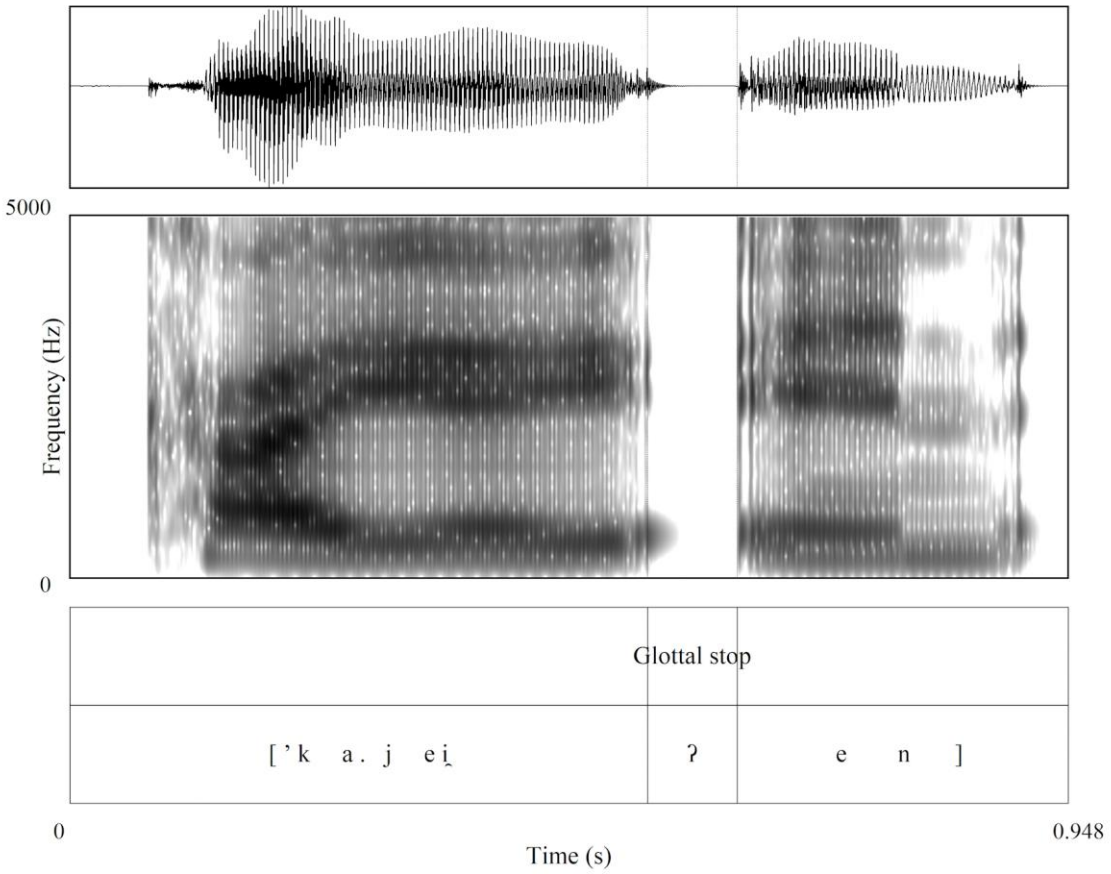

Figure 3: Glottal stop (80 ms closure) between two vowels in the word sequence *cae en* 'he falls in'[9]

4.2 Linked versus separated sequences: binary distribution

The frequencies and distributions of the resolution types are given in Table 2. The results indicate that monolinguals do not categorically produce continuous sequences but instead produce a variable proportion of discontinuous tokens (7%) separated by a glottal stop and, in a few cases, consonantal insertion. These findings for monolingual speakers indicate relatively high rates of vowel linking, which is in line with previous research by Smith/Flores/Gradoville (2008).[10] Thus, there appears to be an increase in native-like behavior as target language proficiency rises. All groups are significantly different based on rates of separation, χ^2 (2) =231.24, p <2.2e-16. Multiple comparisons for the separated condition showed significant differences.

[9] We included a phonetic symbol for the [j] to represent the high-front vocoid. Although the phone is not part of the lexical representation, the speaker inserts the sound so it is included in the transcription in addition to the gliding of the /e/.

[10] We note that Smith/Flores/Gradoville (2008) primarily differentiated between hiatus and diphthong on formant movements evaluated at three points in the sequence. Sequence duration was included as a additional variable and they report that the duration was typically longer in sequences with a hiatus.

Group	Linked #/%	Separated #/%	Total
Monolinguals	260/93%	19/7%	279
Advanced (Graduate students)	588/71%	240/29%	828
Intermediate (Undergraduates)	271/32%	571/68%	842
Total	1119	830	1949

Table 2: Linked versus separated vowel sequence resolutions (raw and percentage) by speaker group

Figure 4 provides a Mosaic plot of the distribution of tokens as a percentage. The x-axis contains the three speaker groups, while the y-axis represents the two resolution strategies in a binary distinction of linked versus separated. The image captures the two extremes of monolinguals with high levels of linking versus the undergraduate intermediate learners with less than a third of their productions realized as linked vowels. The advanced learners or second language group comprised of graduate students demonstrate an evident change toward the use of linked speech normally associated with monolingual speakers.

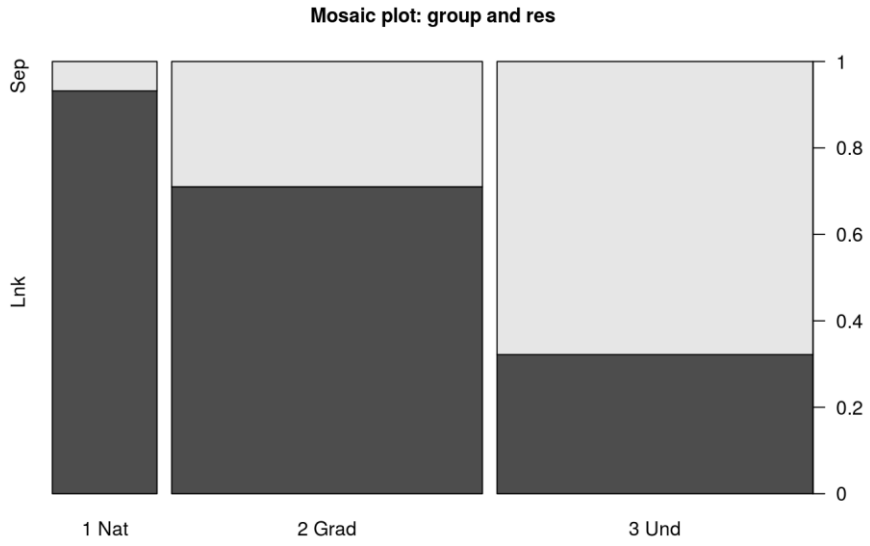

Figure 4: A mosaic plot of speaker groups and vowel sequence of linked versus separated resolutions across word boundaries (1. Nat=Monolingual, 2. Grad=Advanced, 3. Und=Intermediate)

4.3 Resolution patterns of linked versus separated categories

In the linked resolution category, there were several types or realizations of vowel linking. Likewise, there were several types within the separated category. Examples are presented in the following sections.

4.3.1 Linked vowel sequences

We used the following guidelines for sequence resolution based on segmental duration from previous literature for singleton vowels, observations from our own perceptual characterization, and the presence of steady-state formants. Sample patterns are presented in table 3.

Merged	le_hace ‘does to him’	[ˈlɛː.se]	
Replaced	lo_importante ‘the important thing’	[liːm.poɾ.ˈtaɲ.te]	
Diphthong	su_otro ‘his other’	[ˈsu̯o.tɾo]	
Hiatus		la_otra ‘the other’	[la.ˈo.ɾa]

Table 3: Linked vowel variants

The linked vowels sequence category had several resolution types, including a merged vowel, a replaced vowel, a diphthong, or a hiatus, and were based on the vowel sequence duration and formant patterns. The **merged vowel** characterization involved a single vocoid that was mid-way between the two vowels. The **replaced** vowel designation was when one of the two vowels was omitted, and the other was typically lengthened. The **merged** variant typically had a short duration of a single vowel based on the formants and often was due to both vowels being of the same phonemic quality. The **replaced** variant also had a short duration commensurate with a singleton vowel with formant values associated with one of the vowels in the vowel sequence. The duration for both the **merged** and **replaced** variants was generally <100 ms.

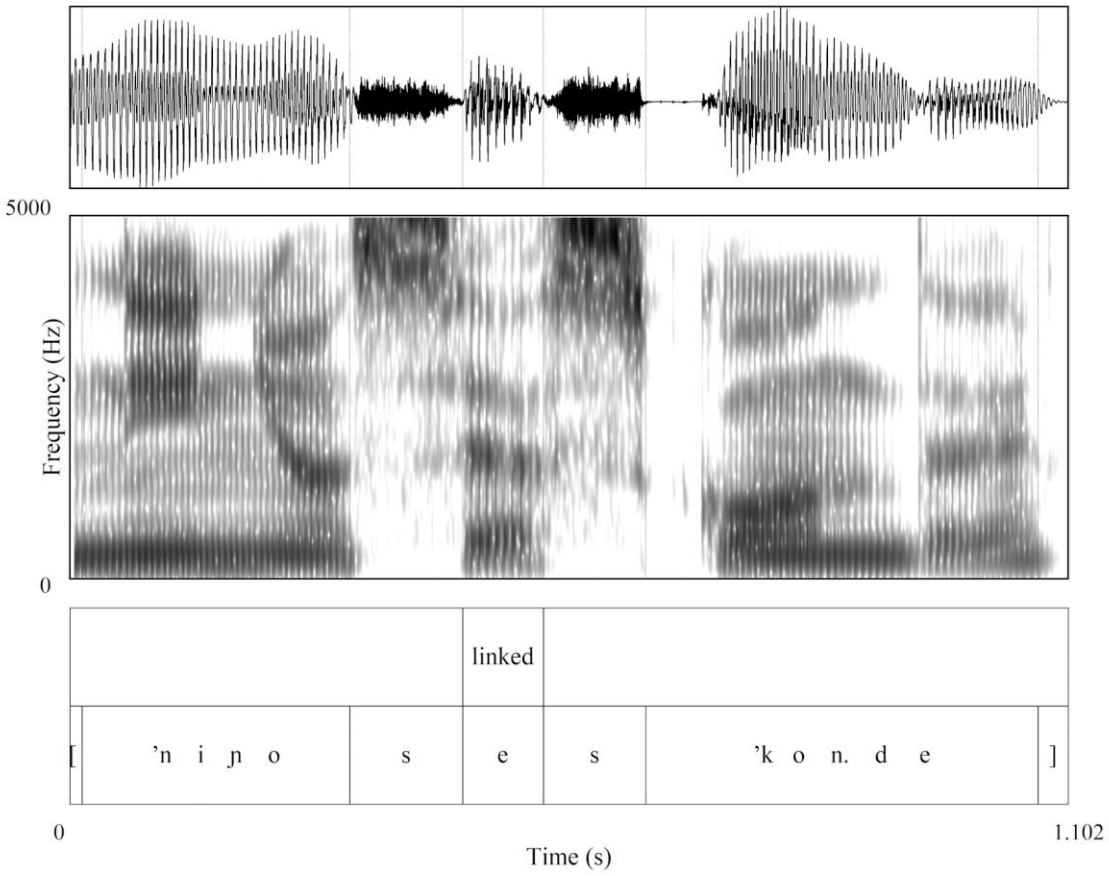

Figure 5: Linked vowels (89 ms) in the word sequence *niño se esconde* ‘boy child hides’

The diphthong variety implies a glide with a steady-state vowel sequence and could occur in either order. The **diphthong** variant typically had a clear, steady-state vowel of at least 50 ms with a glide portion of variable duration. The durational range for the entire diphthong variant varied between 80–160 ms. Figure 6 presents an example of a diphthong with a longer duration (146 ms). The lack of an initial steady-state vowel at the sequence onset is the determining factor in categorizing this production as a diphthong. We also note that the F1 value was 432, and the F2 value at the onset was 2154 Hz, which is in contrast to the reported values for front vowels in Spanish (cf. Quilis/Esgueva 1983).

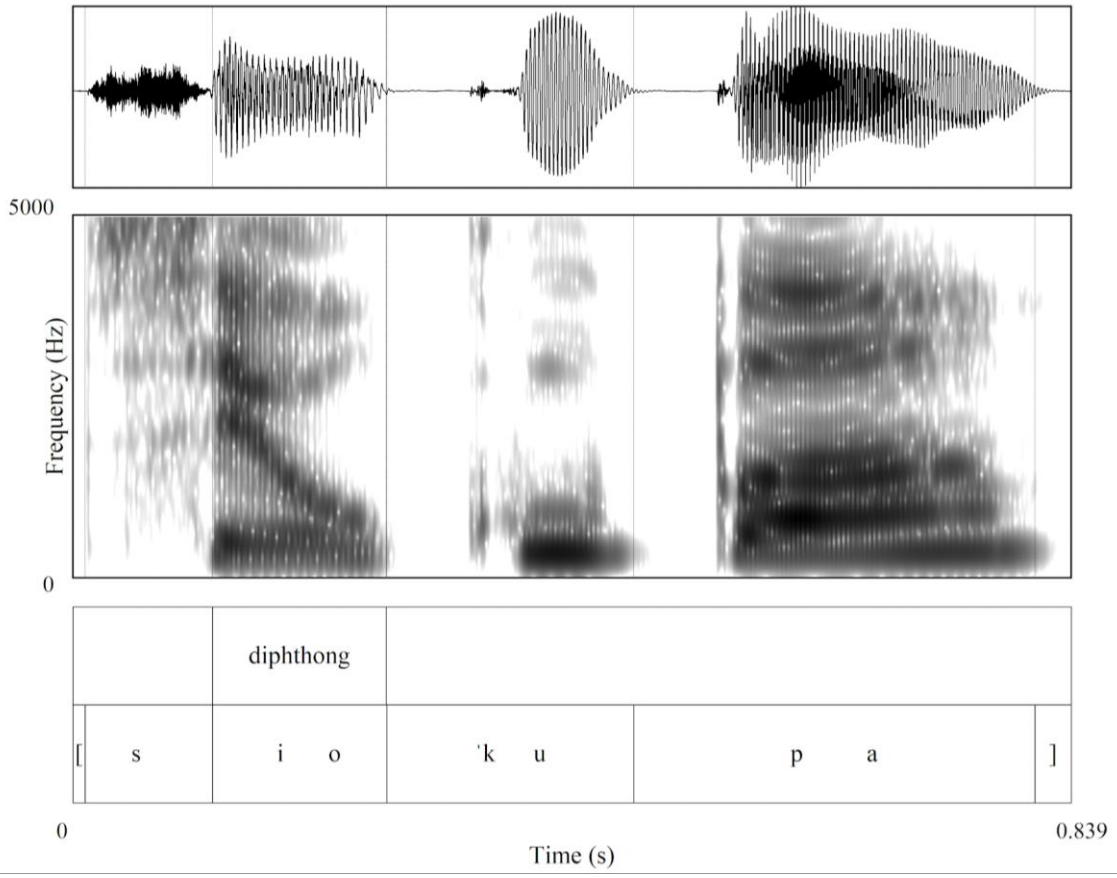

Figure 6: Spectrogram of an example of a two-vowel sequence as a longer diphthong (146 ms) in the word sequence *se ocupa* 'he occupies'

The final resolution type was a hiatus (separate vowels belonging to distinct syllables), and we would observe two relatively steady states of formant values. The **hiatus** variant was produced with two distinct steady-state formant configurations with a combined duration above 160 ms.

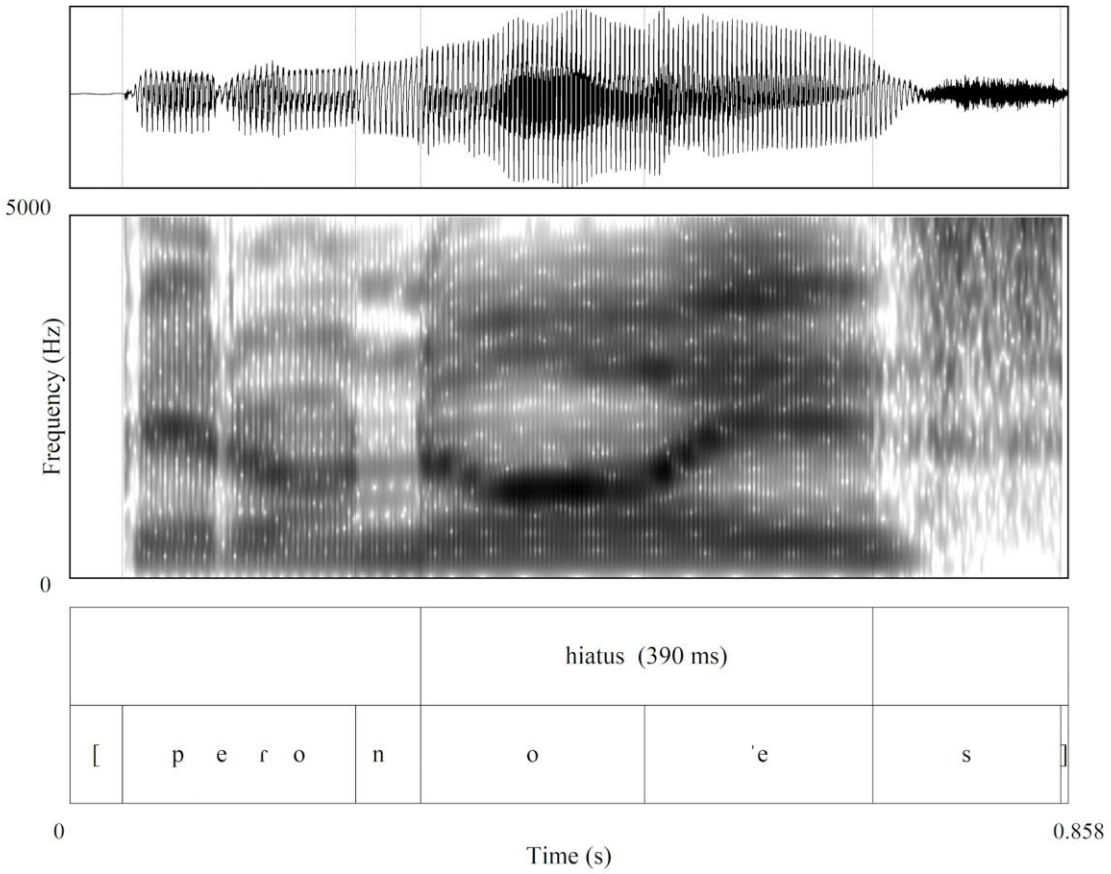

Figure 7: Spectrogram of a two-vowel sequence with vowel hiatus (390 ms) in the word sequence *pero no es* 'but it is not'

Figure 7 shows a long duration for the vowel sequence with a clear steady-state formant structure for each of the vowels and a smooth transition between them. This longer duration hiatus was more common among the second language speakers without creaky voice or a glottal stop seen in the separated tokens.

Speaker groups	Total tokens	Linked category tokens	Merged/replaced[11]	Diphthong	Hiatus
Monolingual	279	264/94%	147/53%	91/33%	26/9%
Advanced	828	588/71%	171/21%	276/33%	141/17%
Intermediate	842	271/32%	55/7%	54/6%	162/19%
	1949	1,123	373	421	329

Table 4: Linked/continuous vowel sequence patterns

[11] The merged and replaced classification types were collapsed for several reasons. The primary reason is that they reflected a single vowel production with a short duration due to being the same vowel or being produced as a single merged vowel, or single glide-vowel sequence without any steady state of the vowel formants (the entire vowel sequence was a transition, and very short). The second reason was due to the lack of distinct formant structures that would indicate two distinct vowels.

Table 4 presents the tokens and frequency of productions in raw counts (total tokens produced) and percentages based on the entire token corpus to facilitate comparisons across the linked and separated categories. The intermediate learners linked the vowel sequences most often with hiatus, slightly more frequently than the advanced learners. However, the advanced learners produced considerably more merged and replaced tokens than the intermediate learners. The monolingual speakers produced the highest number of merged/replaced tokens at 53%, followed by diphthongs at 33%, and hiatus was the least common linked production. The merged and replaced category showed the clearest distinction in patterns between levels, with the monolinguals producing over 50% down to 21% by the advanced learners, with the intermediate learners producing only 7%, indicating a large gap between the intermediate learners and monolinguals.

4.3.2 Separated vowel sequences

As mentioned previously, there were three types of strategies for the separated category of resolution, and the most frequent resolution was a glottal stop that often had several pulses of glottal creak on the periphery, as exemplified in Figure 3.

The intermediate learners' second most common separation strategy was to insert a pause. In Figure 8, the pause between the two vowels is 222 milliseconds. The duration of the pause ranged between 75 milliseconds and 300 milliseconds. Cases with a separation that exceeded 300 milliseconds were not included as they were no longer a running speech sequence but reflected a pronunciation or lexical recovery challenge.[12]

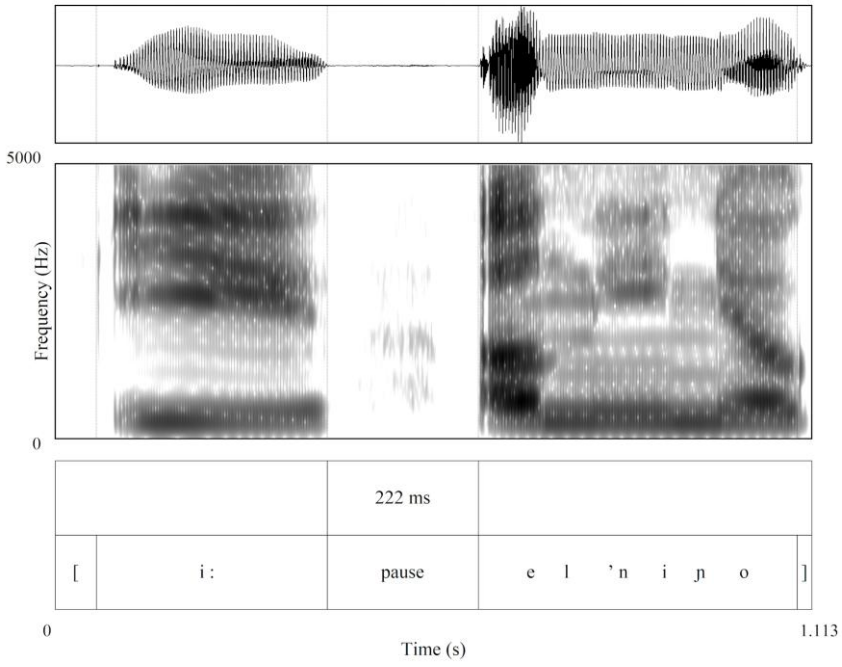

Figure 8: Spectrogram of the separated category of pause (222 ms) between two vowels in the word sequence *y el niño* 'and the boy'

[12] This 300 ms choice is arbitrary and reflected our sense that the speaker did not know what to say next as opposed to a specific desire to separate vowels.

The resolution type of creaky voice between the two vowel sequences (see Figure 9) was the least common resolution for all speakers, including the intermediate speakers; however, the frequency still reached almost a fifth of the total intermediate-level productions. In some cases, the glottal stop and creaky voice categories demonstrated a degree of overlap.

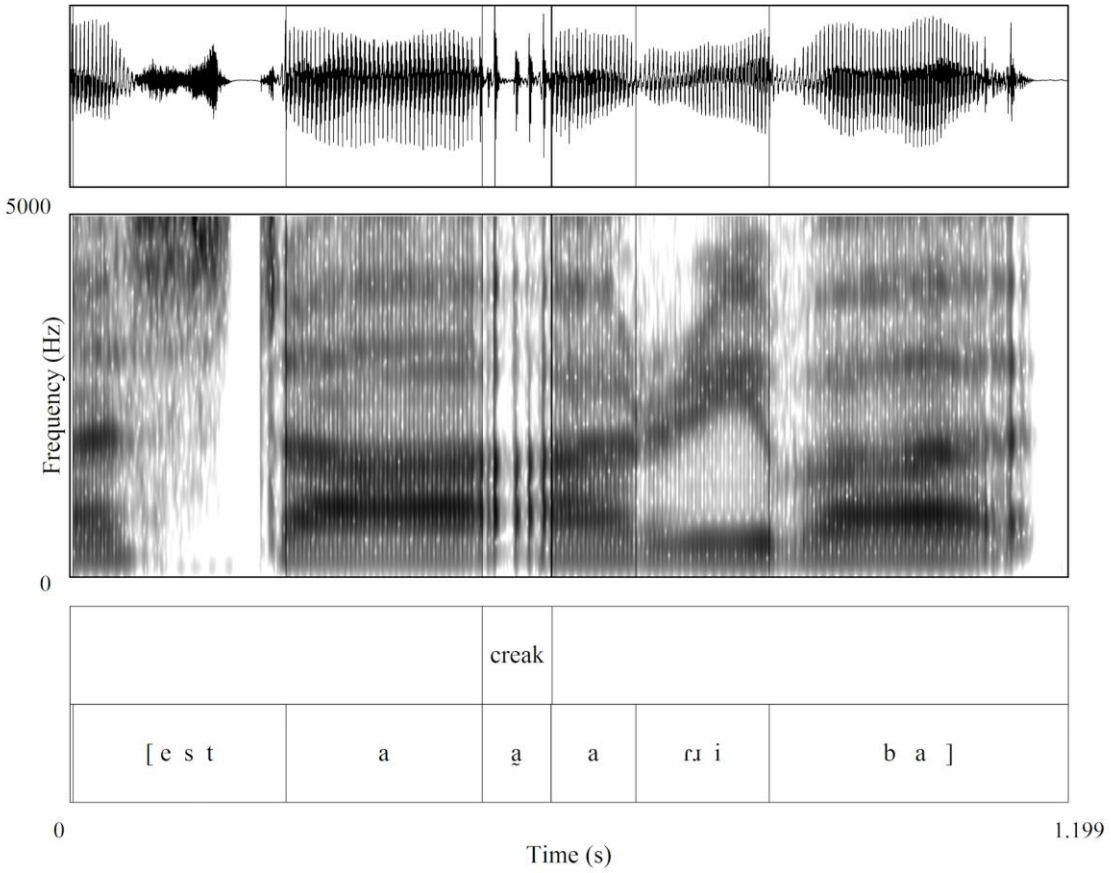

Figure 9: Creaky voice or glottal fry between two vowels (417 ms with 87 ms of creak) in the word sequence *estaba arriba* 'he was above'

Figure 9 is another example of an identical vowel sequence that prescriptively should be produced as a single vowel. The two vowels have the same formant structure, yet there is a clear separation due to the creak of the vowel sequence. Figure 10 is another example of separated distinct vowels with a clear transition between the mid vowels /e/ and /o/, and the transition is marked with creaky voice.

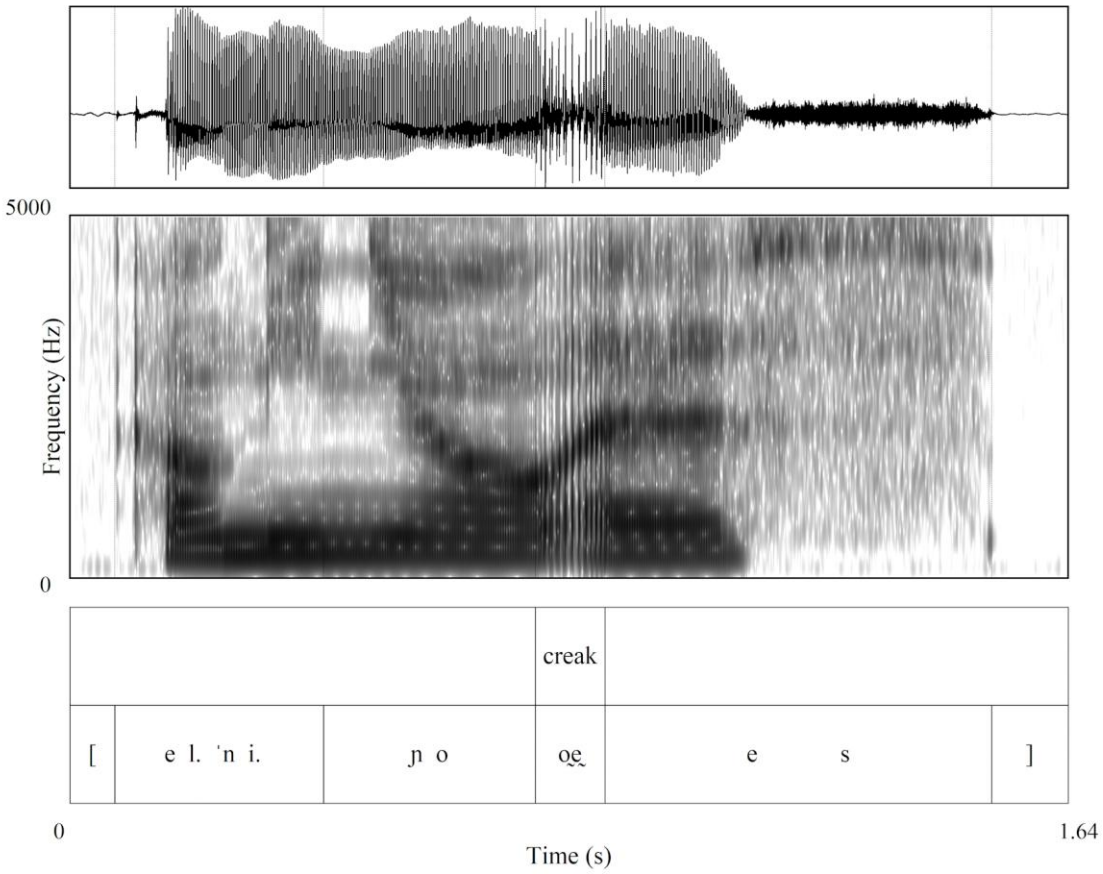

Figure 10: Word boundary vowels with creaky-voice transition (114 ms) in the word sequence *el niño es* 'the boy is'

Table 5 presents the cross-tabulated data for the resolution types by speaker groups. The first number is the raw token count, followed by the percentage. The token count and percentages included reflect the percentages of overall tokens analyzed, not just the separated resolution type. We chose to present the percentages in this manner to facilitate comparisons across the resolution categories and types.

Speaker groups	Total tokens	Separated category	Glottal	Pause	Creak
Monolingual	279	15	6/2%	7/3%	2/1%
Advanced	828	240	95/11%	108/13%	37/4%
Intermediate	842	571	220/26%	192/23%	159/19%
	1949	826	321	307	198

Table 5: Separated/discontinuous vowel sequence resolution types

4.4 Task types

This section documents the resolution categories **linked** and **separated** across the two learner groups based on the task. The monolingual speakers only produced the narrative retelling, and their data is included for comparison. In Figure 11, each resolution category is presented as a percentage of the tokens produced in that task to facilitate the comparison across tasks.

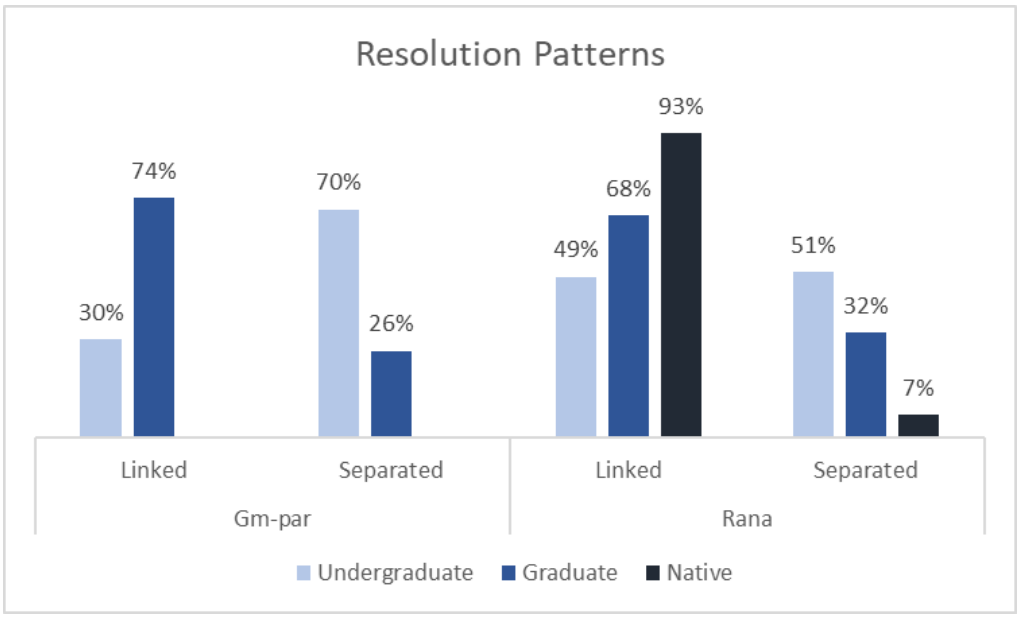

Figure 11: Resolution patterns by task type (Gm-par = reading task; Rana = narrative retell)

The intermediate learners only had 30% linked vowels in the reading task as compared to 49% in the picture narrative. Viewed across task types, the intermediate learners demonstrate a 40% difference spread towards separated vowels in the reading task, whereas there is only a 2% difference in the picture narrative. The advanced learners demonstrated a clear preference for linked vowels in both tasks, with a slightly higher rate of linking in the reading task, 74% compared to 68%. Across all three speaker groups, as seen previously, the monolingual speakers have the highest frequency of linked vowel sequences; however, there is some variability in their resolutions. For the learner groups, a Chi-squared test with Yates' continuity correction showed significant differences between the two task types for the two L2 speaker groups for the production of separated vowel sequences (χ^2 (1) = 8.014, p-value = 0.004642).

5 Discussion

5.1 Research questions

Our first research question was, "How do English L1-speaking learners of Spanish resolve contiguous vowels across word boundaries compared to monolingual speakers?" The two learner groups produced the vowel sequences, both linked and separated resolutions, with variations in their rates of usage. Intermediate learners demonstrate separated or discontinuous vowel sequences with high rates of separated vowels using several strategies, such as glottal stops, pauses, and creaky voice, that are common in American English phonological patterns. These same separation strategies were found among the advanced learners, though at much lower rates of usage.

The second research question sought to determine if there were particular patterns or resolution strategies among the groups for both linking/continuous and separated/discontinuous sequences. The data results suggest that the patterns found among the intermediate speakers also occur with the advanced speakers at a lesser rate of application. In the linked vowel category, the monolingual speakers produced higher rates of reduced vowels with shorter overall dura-

tions. The intermediate speakers tended to produce vowel sequences with a longer duration, which will be explored elsewhere. The advanced learners produced linked resolution types with durations that fell in the middle between the intermediate speakers and the monolinguals. The duration of vowel sequences seems to function as a cue to monolingual-like pronunciation; the shorter the sequence duration, the more similar the monolingual speech patterns are. In the separated vowels sequence, the monolinguals only produced two tokens of creaky voice, reflective of 1%. The intermediate learners produced more glottal stops than the advanced learners, who produced slightly more pauses than glottal stops. Again, rates of separation were much lower for the advanced learners.

The third research question was concerned with the role of task type in the resolution of vowel sequences. The intermediate learners produced more linked vowel sequences in the narrative retelling compared to the textbook reading. The advanced learners produced slightly more linked vowel sequences in the reading task. The reading task does not require the speaker to retrieve vocabulary or create novel utterances. The lack of required planning for communication could encourage a more fluid pronunciation, which is what happened with the advanced learners. Interestingly, the intermediate learners were more likely to produce linked vowel sequences in the narrative retelling and far less in the reading task. Despite the differences in rates of linking and separation between the two learner groups and the two tasks, it is important to reiterate that the advanced learners produced considerably more linked vowel sequences in both task types than the intermediate learners did. A question arises as to why the context with greater cognitive load, which requires lexical retrieval and utterance creation, would facilitate an increase in monolingual Spanish-linked speech patterns for the learners.

The current study did not examine the lexical density and complexity of the narratives between the two learner groups. However, this may be a fruitful line of research for future studies.

5.2 Linking continuum

The data suggest a continuum of contiguous vowel resolutions, as illustrated in Figure 12. In addition to the descriptive qualities of linking or separation, the duration of the sequences also plays a role, as previously described. The durations of the linked categories of diphthong and hiatus slightly overlap, and category selection was based on formant behavior. A hiatus could have a shorter duration when two steady-state vowels were present, based on the formants. In contrast, the diphthong category had only one steady-state vowel with a glide, which was occasionally produced with a longer duration. A similar overlap occurred between the merged/replaced and diphthong types, which were identified based on formant behavior.

We acknowledge that these resolution characterizations and descriptions involve a degree of subjectivity. However, we suggest that duration measurements and formant criteria offer a more nuanced and replicable approach to characterizing and classifying contiguous vowel production across word boundaries than previous studies.

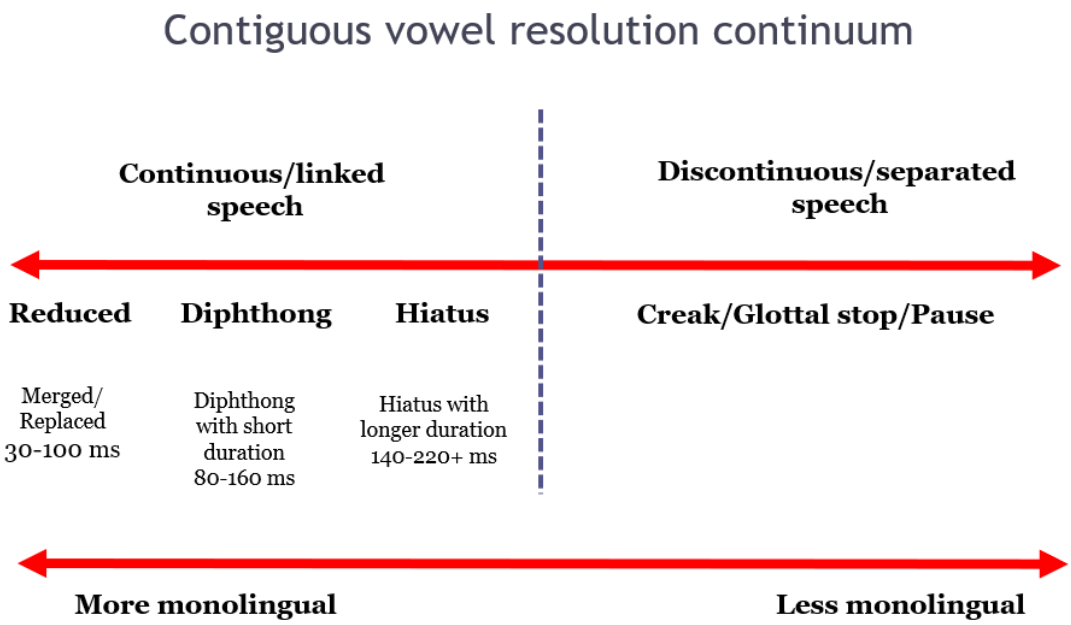

Figure 12: Continuum of vowel # vowel resolution strategies

5.3 Tasks and Learners

There seems to be a tension between the formality of the texts and the nature of the speech produced for each, which relates to formality and speech planning. At the same time, the reading task about a Nobel Prize-winning Colombian Spanish author could be considered more formal due to the more advanced syntactic structure and lexical density compared to a child's simple illustrated picture book. The picture book required the participants to create a coherent narrative that followed the illustrations, including the vocabulary and syntactic structure. One possibility is that the advanced learners, graduate students, were very comfortable reading about Spanish literary authors and could focus their speech on producing a fluid and presentational manner of speaking. Additionally, the advanced learners would have read and discussed this type of passage regularly in their graduate coursework, and they likely functioned as the instructors of this or a similar passage to other intermediate-level learners. While the intermediate learners would have read this exact passage or passages very similar to it, the formal reading passage contained vocabulary and syntactic structures that were not typical to their usage of Spanish. The intermediate learners may have benefited from their decision to use frequently occurring lexemes and word combinations that are familiar, which could result in the more fluid pronunciation and linking of contiguous vowels across word boundaries.

5.4 Acquisition of a process

There has been considerable research on the acquisition of new sounds in language acquisition and how to model or theorize the acquisition process. The current phenomenon of synalepha is not concerned with new sounds in the acquired language, which in this case would be a simplification of the vowel system and avoidance of schwa, but rather with the interaction of sounds. In other words, the sounds themselves can be the same, but how they interact or overlap, in this case, becomes the phenomenon that must be acquired.

This linking of sounds across word boundaries does occur in very informal speech in American English, often in conjunction with additional consonantal reduction. If this linking of vowels across word boundaries functions as a reductive or lenitive process, then the acquisition of synalepha could be quasi-compared to other allophonic variants in the acquisition of Spanish, such as voiced stops as approximate allophones in intervocalic position.

The previous section showed variable linking patterns for different tasks. However, simpler reading tasks could potentially allow learners to practice linking vowels. Additionally, high-frequency word pairs could be practiced to entrench or solidify the linking patterns that often occur with high-frequency lexemes and lexical pairs. The results suggest that narrative retells yield more nativelike vowel linking among intermediate learners and may be a worthwhile task to explore further and incorporate into classroom practices.

6 Conclusion

The study explores the production of contiguous vowels across word boundaries by two groups of L1 English speakers learning Spanish compared to monolingual Spanish speakers. All three groups employed similar production types but with different rates. Intermediate learners frequently separated vowel sequences with glottal stops, pauses, and creaky voice, common in American English. Advanced learners used these separation strategies less frequently, showing a progression toward monolingual Spanish patterns. Monolingual speakers typically produced continuous speech with shorter, linked vowel sequences and with shorter durations. The advanced learners produced vowels with durations that fell between those of the intermediate learners and monolinguals. The study highlighted that the duration of vowel sequences serves as a cue to more monolingual-like pronunciation.

Task type also played a role in vowel sequence resolution, with intermediate learners producing more linked vowels in narrative tasks compared to reading tasks. Advanced learners showed the opposite trend, potentially due to their familiarity with formal texts and the cognitive load associated with formal texts. The advanced learners' ability to produce more linked vowel sequences in both task types than intermediate learners suggests that context and task complexity influence pronunciation patterns variably. Future research could explore lexical density and complexity in narratives to further understand their impact on vowel sequence resolution. Ongoing research by the authors explores linguistic variables with acoustic details. Overall, the study provides an initial and fundamental descriptive documentation of the process of contiguous vowel resolutions by multiple speaker groups, important acoustic details of the contiguous vowel resolutions, and beneficial insights into the phonological adjustments learners make as they advance in their Spanish proficiency.

References

Alba, Matthew (2005): *Hiatus resolution between words in New Mexican Spanish: A usage-based account*. Ph.D. dissertation, University of New Mexico.

Alba, Matthew (2006): "Accounting for variability in the production of Spanish vowel sequences." In: Sagarra, Nuria/Toribio, Almeida Jacqueline (eds.): *Selected proceedings of the 9th Hispanic Linguistics Symposium*. Somerville (Massachusetts), Cascadilla Proceedings Project: 273–285.

Alba, Matthew (2008): "Ratio Frequency: Insights into Usage Effects on Phonological Structure from Hiatus Resolution in New Mexican Spanish." *Studies in Hispanic and Lusophone Linguistics* 1: 247–286.

Barbería Aurrekoetxea, Irene (2012): *The gradient nature of hiatus resolution in Spanish*. Dissertation, Universidad de Deusto, San Sebastián.

Beristain, Ander (2024): "The Role of (Re)Syllabification on Coarticulatory Nasalization: Aerodynamic Evidence from Spanish." *Languages* 9: 219.

Boersma, Paul/Weenink, David (2021): *Praat: doing phonetics by computer*. praat.org [12.12.2021].

Bowen, J. Donald/Stockwell, Robert P. (1960): *Patterns of Spanish Pronunciation: A Drill-book*. Chicago: University of Chicago Press.

Chitoran, Ioana/Hualde, José Ignacio (2007): "From hiatus to diphthong: The evolution of vowel sequences in Romance." *Phonology* 24/01: 37–75.

Colina, Sonia (2009): *Spanish Phonology: A Syllabic Perspective*. Washington, D. C.: Georgetown University Press.

Davidson, Lisa/Erker, Daniel (2014): "Hiatus resolution in American English: The case against glide insertion." *Language* 90/2: 482–514.

Face, Timothy/Alvord, Scott (2004): "Lexical and Acoustic Factors in the Perception of the Spanish Diphthong vs. Hiatus Contrast." *Hispania* 87: 553.

Geeslin, Kimberly L. (2011): "Variation in L2 Spanish: The state of the discipline." *Studies in Hispanic and Lusophone Linguistics* 4/2: 461–518.

Geeslin, Kimberly L. (2003): "A comparison of copula choice: Native Spanish speakers and advanced learners." *Language Learning* 53/4: 703–764.

González, Carolina/Weissglass, Christine (2015): "Task effects in resyllabification: L1 and L2 differences." Paper presented at the *34th Annual Second Language Research Forum*, October 29–31. Atlanta, Georgia.

Harris, James W. (1983): *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.

Holt, David Eric/Gelormini-Lezama, Carlos (2008): "On the acquisition of synalepha and resyllabification in Spanish by English-speaking advanced learners." Paper presented at the *Current Approaches to Spanish and Portuguese L2 Phonology Conference*, February 22–23. University of Minnesota.

Hualde, José Ignacio/Prieto, Mónica (2002): "On the diphthong/hiatus contrast in Spanish: Some experimental results." *Linguistics* 40/2: 217–234.

Jenkins, Devin (1999): *Hiatus resolution in Spanish: Phonetic aspects and phonological implications from Northern New Mexican data*. Dissertation, University of New Mexico, Albuquerque.

Jiménez-Bravo, Miguel/Lahoz-Bengoechea, José María (2023): "Durational cues to resyllabification in Spanish." *Loquens* 10, no. 1–2: e099. doi.org/10.3989/loquens.2023.e099.

Marchini, Gilly/Ramsammy, Michael (2023): "/s/-lenition and resyllabification in Southern Cone Spanishes." *20th International Conference of Phonetic Sciences* (ICPhS). Prague, Guarant International: 2144–2148.

Martínez-Gil, Fernando (2000): “La estructura prosódica y la especificación vocálica en español: El problema de la sinalefa en ciertas variedades de la lengua coloquial contemporánea.” In: Fernández, Juana Gil. (ed.): *Panorama de la fonología española actual*. Madrid, Arco Libros, S. L.: 511–560.

Mayer, Mercer (1969): *Frog, Where Are You?* New York: Dial Press.

Morris, Richard E. (2010): *Pronunciación de la lengua española para anglohablantes*. Indianapolis: Focus.

Olive, Joseph P./Greenwood, Alice/Coleman, John (1993): *Acoustics of American English Speech: A Dynamic Approach*. New York: Springer.

Quilis, Antonio (1993): *Tratado de fonología y fonética españolas*. Madrid: Arco.

Quilis, Antonio/Esgueva, Manuel (1983): “Fonemas vocálicos españoles.” In: Esgueva, Manuel/Cantarero, Margarita (eds.): *Estudios de fonética*. Consejo Superior de Investigaciones Madrid: Científicas, Instituto Miguel de Cervantes.

Real Academia Española (2011): *Nueva gramática de la lengua española: Fonética y fonología*. Madrid: Espasa.

Souza, Benjamin J. (2010): *Hiatus resolution in Spanish: An experimental study*. Ph.D. dissertation, The Pennsylvania State University.

Smith, Janet M./Flores, Tanya L./Gradoville, Michael S. (2008): “An analysis of vowels across word boundaries in Veracruz, Mexican Spanish.” IULC *Working Papers* 8/1.

Stockwell, Robert P./Bowen, J. Donald/Martin, John W. (1965): *The grammatical structures of English and Spanish*. Chicago: University of Chicago Press.

Whitley, Melvin Stanley (2002): *Spanish/English contrasts: A course in Spanish linguistics*. Washington, D.C.: Georgetown University Press.

Willis, Erik W. (2008): *Corpus of Veracruz, Spoken Spanish: Frog stories*. Private corpus.