Too Close: Predicting Segmentation Behaviors of Beginning Readers from Coarticulatory Patterns in English and Spanish

By

Himilcon Inciarte

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor A. Cunningham Carson, Chair

Professor S. Rabe-Hesketh

Professor S. Gahl

Spring 2026

Too Close: Predicting Segmentation Behaviors of Beginning Readers from Coarticulatory Patterns in English and Spanish

By

Himilcon Inciarte

Abstract

Too Close: Predicting Segmentation Behaviors of Beginning Readers from Coarticulatory Patterns in English and Spanish

by

Himilcon Inciarte

Doctor of Philosophy in Education

University of California, Berkeley

Professor A. Cunningham Carson, Chair

This dissertation tests the premise that, for beginning readers, speech is the 'object of segmentation' (Schreuder and van Bon, 1989). I posit that one aspect of speech in particular—coarticulation as proxied by gestural overlap—is likely to influence segmentation performance: reduced coarticulation should enable segmentation, while greater overlap should make it more difficult. Using empirical evidence of gestural overlap from articulatory Phonetics, I make predictions about the segmentability of English and Spanish stimuli, both between and within languages.

Participants were 472 TK – Grade 2 children from various educational contexts. Students were asked to segment a spoken word. Their responses received two scores: first sound isolation (0/1) and unitization (0/1), the latter denoting instances where the first two phonemes of a stimulus were identified as a single unit (e.g., /fl/ for flee). For each outcome, 12,388 responses were analyzed within a Generalized Linear Mixed Model (GLMM) framework that controlled for item- and person-side variables.

Five out of six hypotheses were supported, pointing to a strong, predictable association between gestural overlap and segmentation performance. Specifically: 1) English clusters were more difficult to segment than Spanish clusters; 2) English items starting with vowels followed by sonorants (e.g., eel) were more difficult than their Spanish counterparts and English equivalents with obstruent codas (e.g., eke); 3) English and Spanish singleton onsets did not differ in difficulty; 4) voiced-plosive clusters were more difficult than unvoiced-plosive clusters in both languages; 5) fricative clusters were easier to segment than plosive clusters in both languages.

Taken together, these results suggest that the speech code is indeed what participants operate on during a phoneme segmentation task. These findings are discussed within the context of broader

literature on theory-development, reading instruction, and assessment.

Keywords: Phoneme Awareness; Coarticulation; Gestural Overlap; GLMM; Explanatory IRT; Spanish/English Bilinguals

Table of Contents

List of Tables

List of Figures

Dedication

To *Mamá* and *Papá*, who left behind everything they knew so that I could dream.

And to Rachael, Lucky, Leesán, and Lorel, who uprooted their lives in the pursuit of my dream.

Acknowledgements

I would not have been able to complete this degree without all the support I had along the way.

My chair, Anne Cunningham: Your steadfast guidance has been invaluable and unparalleled. The clarity of your feedback and suggested next steps; your focus on progress without sacrificing quality; and your ongoing encouragement, especially at times of personal and professional difficulty. I could not have finished this degree without you. Thank you.

Sophia Rabe-Hesketh: Your ability to explain complex statistical concepts—often multiple times—without ever making me feel like I was somehow unqualified is something I will always treasure. The models employed in this dissertation, and my ability to interpret them, are a direct result of your mentorship.

Susanne Gahl: Your eagerness to support this journey despite my lack of linguistics training speaks to your commitment and kindness. I especially appreciate how you have pushed me to learn and improve instead of letting me settle.

Mark Wilson: You are the reason I came to Berkeley, and I am so grateful for all that you have taught me.

To all the administrators, teachers, and families of the students who participated in my studies: Thank you for your trust in me.

And finally: Thank you to all my peers from Anne's, Sophia's, and Mark's research groups for keeping me grounded and being invaluable sounding boards.

Introduction

Learning to segment speech into smaller phonological units is associated with later word reading success (Melby-Lervåg et al., 2012; Míguez-Álvarez, 2021). Beginning readers might struggle to develop this insight, particularly as it pertains to phonemes (e.g., Cunningham, 1990; Blachman, 1994; Erbeli et al., 2024; Rehfeld et al., 2022). Difficulty with phoneme awareness (PA) is not surprising given that spoken language is not merely the utterance of individual sounds in rapid succession (Liberman et al., 1967). Therefore, segmentation cannot simply mean identifying beginning and end points of sounds corresponding to underlying phonemes in the acoustic signal (Ehri et al., 1987; Skjelfjord, 1987b).

What information, then, does a beginner reader leverage to segment words into phonemes? Schreuder and van Bon (1989) proposed that "… an articulatory code is the object of segmentation" (p. 75). This would suggest that they rely on speech itself. Similarly, Content et al. (1986) used the term "phonetic segmentation" (p. 51) to underscore the task's reliance on articulatory information. Throughout the years, studies have tried to find evidence of this link by using phonological attributes of phonemes to predict segmentation performance (e.g., De Graaff et al., 2011; Geudens & Sandra, 2003; Geudens et al., 2004; McBridge-Chang, 1995), but results have been mixed and difficult to parse.

Against this backdrop, the present dissertation investigates the relationship between a specific aspect of the articulatory code—coarticulation vis-à-vis gestural overlap—and segmentation performance. The logic undergirding the experiment can be summarized as follows: speech is coarticulated, which is bound to make phoneme segmentation difficult. Gestural overlap is one proxy for coarticulation. If the articulatory code informs segmentation, then onsets with a higher degree of gestural overlap should be more difficult and elicit specific segmentation behaviors.

Subsequent chapters flesh out this logic. First, the literature review (Chapter 1), which has three objectives:

1. summarize early theories linking coarticulated speech to segmentation difficulty;
2. underscore key limitations of subsequent investigations into this phenomenon;
3. introduce the concept of gestural overlap as a proxy for coarticulation;

In Chapter 2, I outline an experiment that investigates whether different articulatory patterns in English and Spanish predict differences in segmentation performance. Finally, I describe the methods employed for data collection and analysis (Chapter 3). Chapter 4 presents the experiment's results, and Chapter 5 discusses these findings.

Chapter 1
Literature Review

This chapter serves three purposes. First, I summarize incipient hypotheses linking speech and segmentation. I then describe the limitations of subsequent PA research exploring this relationship. Finally, I make the case that gestural overlap as operationalized in articulatory Phonetics is worth exploring as a proxy for coarticulation.

1.1 Linking Speech and Segmentation

In this section, I highlight important findings regarding the articulated nature of speech and summarize two early PA studies that illustrate the field's emerging interest in linking this phenomenon to segmentation performance.

Recapping their pioneering work from prior decades (e.g., Liberman et al., 1967; Liberman et al., 1974; Liberman et al., 1977), Liberman et al. (1980) describe the difficulty of matching abstract representations (i.e., phonemes) to the acoustic signal. Using the word 'dog' as an example, they explain: "the information for the three phonological segments is there, but so thoroughly overlapped (encoded) in the sound that there is no way to divide the sound into segments so that each acoustic segment carries information about only one phonetic segment" (p. 140). For that reason, they argue that "awareness of the phoneme … might be difficult for young children" (p. 141) but important to acquire when learning an alphabetic language. These authors, then, view coarticulated speech as an important obstacle to successful phoneme segmentation.

Helfgott (1978) uses that premise as a point of departure in their study of phoneme position as a predictor of segmentation and blending. They write that at least one cause of difficulty with segmentation is "… that in speech, the phonemes are not discrete, linked together like beads on a string, but rather are complexly encoded" (p. 158). This leads them to predict that initial consonants will be more difficult to isolate than final ones since the former are more coarticulated with the vowel. Notably, they offer no empirical evidence for this claim. In the end, data do not support their hypothesis: initial phonemes were significantly easier than final consonants. They interpret this finding as evidence that "initial consonants are more salient, more fully realized, and more clearly articulated than final consonants" (p. 164).

Treiman and Baron (1981) continue this thread in their investigation of phoneme class and PA performance. Of segmentation difficulty, they write, "… that the acoustic representation of a phoneme differs as a function of its context, and that the cues for nearby phonemes are blended together in the speech stream, may account for the difficulty of phoneme awareness" (p. 161). They go on to speculate that vowels and fricatives might have "more invariant" (p. 162) acoustic representations, thereby becoming easier to segment. As with Helfgott (1978), these authors fail to justify their hypotheses—that is, what specifically about fricatives and vowels lead to more stable representations than other phonemes? Once again, results are not in line with predictions: vowels were not easier to segment, and fricatives were easier only in three-phoneme stimuli.

Though not summarized here, other studies throughout the 80s went on to explore a link between speech and PA, including: Content et al., (1986); Ehri (1987); Schreuder and van Bon (1989). All make the same core assumption: since speech is coarticulated, extracting individual sounds from the acoustic signal is challenging. Helfgott (1978) and Treiman and Baron (1981) stand out because their argument goes beyond this initial premise. They contend not simply that coarticulation causes difficulty, but that differences in degree of overlap should cause a commensurate amount of difficulty.

As I show in the next section, the field has since continued this search for gradated differences by relying on phonological variables.

1.2 Phonological Correlates: Sonority and Continuancy

This section argues that two constructs hypothesized to predict PA performance, sonority [± sonorant] and phoneme continuancy [± continuant], are implicitly being treated as proxies for coarticulation.

Sonority denotes the loudness (or vowel-likeness) of a phoneme (Clements, 1990; Parker, 2002). In general, vowels, liquids, and nasals, tend to be categorized as [+sonorant] while fricatives, affricates, and plosives are labeled as obstruents or [-sonorant]. On the one hand, sonority seems like a promising explanatory variable for segmentation since it has well-established roles in phonotactic regularities (e.g., Blevins, 1995; Clements, 1990; Pater, 2009) and phonological processes (for an extensive review, see Parker, 2017). On the other hand, acoustic and phonetic markers of sonority remain elusive (Schröer, 2020).

From an empirical standpoint, prior investigations have led to mixed findings. Several studies have reported that sonorants are more difficult to segment than obstruents (e.g., Yavas & Cogate, 1999; Yavas & Core, 2001); however, these studies did not vary the position of the sonorant phoneme, thereby confounding sonority and position. Even when studies control for phoneme position in the same language, results vary. Consider three studies with Dutch-speaking children. Geudens and Sandra (2003) found that sonorants are more difficult to separate than obstruents in initial (e.g., la vs. fa) and final (e.g., nim vs. fes) positions. On the other hand, De Graaff et al. (2008) reported that only sonorant codas are more difficult (e.g., /l/ in bol vs. lief), a finding replicated by Bouwmeester et al. (2011).

More to the point of this study, any explanation of segmentation performance that relies on sonority suffers from an important limitation: lack of explanatory power. In their comparison of sonorants and obstruents, Geudens et al. (2004) attribute the greater difficulty of sonorants to what they term “phonetic glue” (p. 342). De Graaff et al. (2008) provide a similar explanation, arguing that liquid codas adhere more to the nuclear vowel than obstruents. Crucially, neither study says what this adhesive is. If the claim is that sonorants are more coarticulated with neighboring phones, then data from articulatory studies should be brought to bear (a point acknowledged by Geudens & Sandra, 2003, p. 173).

I now turn to phoneme continuancy, another phonological construct posited to influence PA performance. Continuant phonemes are those that can be stretched without distortion, such as fricatives (e.g., /f/, /s/), nasals (e.g., /m/, /n/), and liquids (e.g., /l/). In contrast, plosives (e.g., /p/,

/t/) and affricates (e.g., /tʃ/, /dʒ/) cannot be elongated without appending a vowel. Among others (e.g., Treiman and Baron, 1981), McBridge-Chang (1995) hypothesized that [+continuant] phonemes might facilitate segmentation because "they can be extended in time" (p. 180). Such an account makes at least two assumptions: (1) children know to direct their attention to the articulatory code; (2) elongation attenuates coarticulation. While there is evidence for the former (e.g., Geudens & Sandra, 2003; Skjelfjord, 1987a), the latter lacks empirical support.

As an illustration of the problem, suppose a child is asked to segment the English word lie. For them to prolong the initial sound, they would need to recognize its essence first (perhaps by way of an articulatory gesture, as in Castiglioni-Spalten & Ehri, 2003). Without that information, it is unclear what the child is elongating. The problem worsens if the stimulus starts with a plosive, such as pie. A child might extend /p/ as /pəː/, not realizing that their production has changed from a consonant to a vowel (i.e., a schwa). While [+continuant] might help disambiguate otherwise confusable phonemes (e.g., /f/ and /s/), it seems that elongation would be most helpful *after* the phoneme has been segmented.

Data from PA studies generally do not lend empirical support to a continuancy advantage. Using English stimuli with third and fourth graders, McBride-Chang (1995) failed to uncover an advantage for [+continuant] consonants, and neither did Treiman and Weatherson (1992). De Graaff et al. (2008) reported similar findings in Dutch with younger children. Moreover, Bouwmeester et al. (2011) found evidence that plosives actually facilitate segmentation, and sonorants (which are generally [+continuant]) had no effect. Jiménez and García (1995) did find evidence in favor of continuants in Spanish. However, their analysis contrasted plosives [-continuant, -sonorant] with a combined category of liquids [+continuant, +sonorant] and fricatives [+continuant, -sonorant], which confounds frication and continuancy.

I draw two conclusions from the above. First, the use of sonority and continuancy as explanatory variables in PA performance have led to mixed results that are difficult to reconcile. Beyond the empirical studies summarized above, consider the following: most sonorant consonants are also continuants (e.g., /m/, /l/). Thus, any proposition that continuancy facilitates segmentation must also account for sonority in some fashion. Some possibilities include: (1) Sonority trumps continuancy (e.g., Geudens & Sandra, 2003); (2) Continuancy trumps sonority (e.g., Jiménez & García, 1995); (3) The effects cancel out; (4) Some other combination. In its present state, the literature cannot arbiter among these scenarios.

Second, and most importantly for the argument being developed here, [+sonorant] and [+continuant] appear to have been employed as proxy variables for coarticulation. In light of the explanations cited in this section, sonorant consonants were believed to be more coproduced with surrounding phones, while continuant phonemes were assumed to be less coarticulated. The challenge is that sonority and continuity are abstract variables created to describe individual phonemes; therefore, they might not necessarily capture articulatory patterns writ large.

My position is that coarticulation is indeed a phenomenon of interest, but that any such investigation should rely on articulatory data.

1.3 Gestural Overlap

This section argues that gestural overlap is an imperfect but promising proxy for coarticulation (a term I use interchangeably with coproduction).

I adopt Browman & Goldstein's (1986) Articulatory Phonology framework, which views gestures as the "phonological primitives" of speech (as summarized in Pouplier, 2020, p. 14). More concretely, gestures are "discrete movements of the speech articulators forming and releasing phonologically relevant constrictions in the vocal tract" (Crouch et al., 2023, p. 1052). It follows, then, that a gesture can be described with regard to place (e.g., whether the lips or tongue tip is involved) *and* time. For instance, producing the phone /p/ involves first, moving both lips together to form a constriction, and then, a release that allows air to burst out.[1]

Combining this premise with Byrd's (1996) definition of coproduction as "the temporal co-occurrence or overlap of two (or more) gestures" (p. 210) yields a blueprint for measuring coarticulation across speech gestures: elapsed time. Gestural overlap, at its essence, is an estimate of how much time has transpired between two gestures. Unsurprisingly, many of its synonyms make this temporal element explicit, including: gestural coordination (e.g., Pouplier et al., 2022), gestural timing (e.g., Byrd, 1996), and temporal coordination (e.g., Loevenbruck et al., 1999). Likewise, point estimates for the lags themselves have gone by various names: IPI (short for interplateau interval, from Gibson et al., 2019 and Sotiropolou et al., 2020); onset lag or plateau lag (both from Marin & Pouplier, 2014 and Pouplier et al., 2022). Nomenclature differences aside, they all attempt to capture the same phenomenon.

Crucially, patterns in temporal overlap have been shown to differ depending on various linguistic factors, such as which phonemes are produced. Chapter 2 summarizes findings relevant to this study in more detail. For now, the important takeaway is that gestural overlap as a construct is poised to capture differences in coarticulatory patterns that can then be used to make testable predictions about segmentation performance.

The remainder of this section describes articulatory data collection and some of its limitations.

Many articulatory studies now use electromagnetic articulography (EMA), a noninvasive method of data collection. Sensors are placed on a predetermined number of articulators, which then allow their movement to be recorded and charted (see Figure 1.3.1, originally from Marin & Pouplier, 2010). One such output is shown in Figure 1.3.2 from Gibson et al. (2019). The y-axis describes the movement of the articulator throughout the gesture, while the x-axis represents time. In addition, superimposed vertical lines (labeled a-f) mark different moments of interest, such as a consonant's release of constriction.

As can be inferred from Figure 1.3.2, a notable challenge for these studies is that human speech is not composed of clearly demarcated phonemes. This means that investigators must make subjective decisions, ranging from seemingly trivial (e.g., exactly where to place sensors) to glaringly important (e.g., exactly when a gesture is released). While these choices are by no

[1] The coupled oscillator model of gestural timing (Nam & Saltzman, 2003; Nam et al., 2009) builds on these gestures to explain how they coordinate to form syllables.

means haphazard, there is not always a single, definitively correct answer (see Rebernik et al., 2021 for more on this topic).

A second challenge is that, despite its intuitive definition, gestural overlap might be operationalized differently across studies. For instance, Gibson et al. (2019) used timestamp e – timestamp b (refer to vertical lines in Figure 1.3.2) to estimate overlap. In such parametrizations, negative values imply more overlap, and positive values mean more time between gestures and hence, less overlap. Conversely, Bombien and Hoole (2013) coded their variables such that positive values indicate more overlap, while negative values denote less overlap (p. 542). These different parametrizations mean that interpreting results from this literature demands careful attention to variable construction.

With these complexities in mind, I clarify that my argument is not that gestural overlap indexes coarticulation perfectly. Rather, I offer that it is a stronger candidate than the phonological constructs described earlier (i.e. the binary features [± sonorant] and [± continuant]). In the next chapter, I outline an investigation using English and Spanish as case studies.

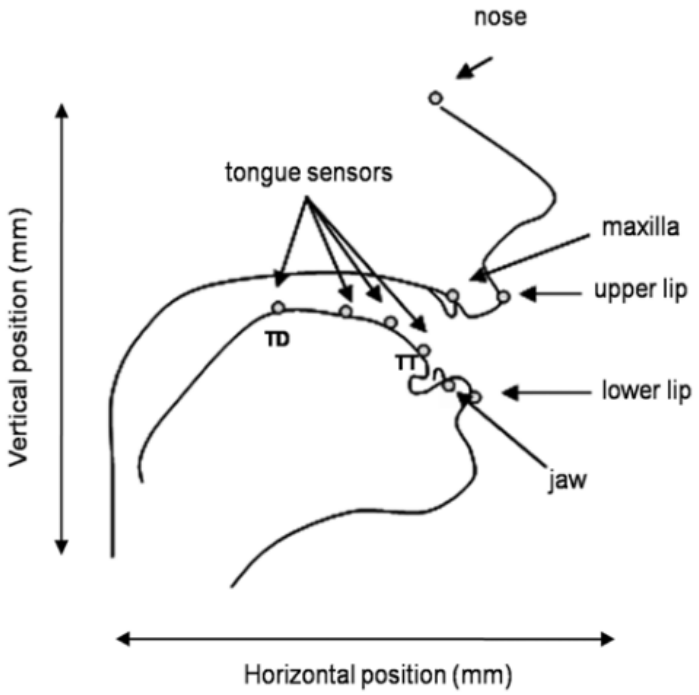

Figure 1.3.1 *Example of sensor placement in EMA study*

TT = tongue tip; TD = tongue dorsum. Reproduced from Marin & Pouplier, 2010.

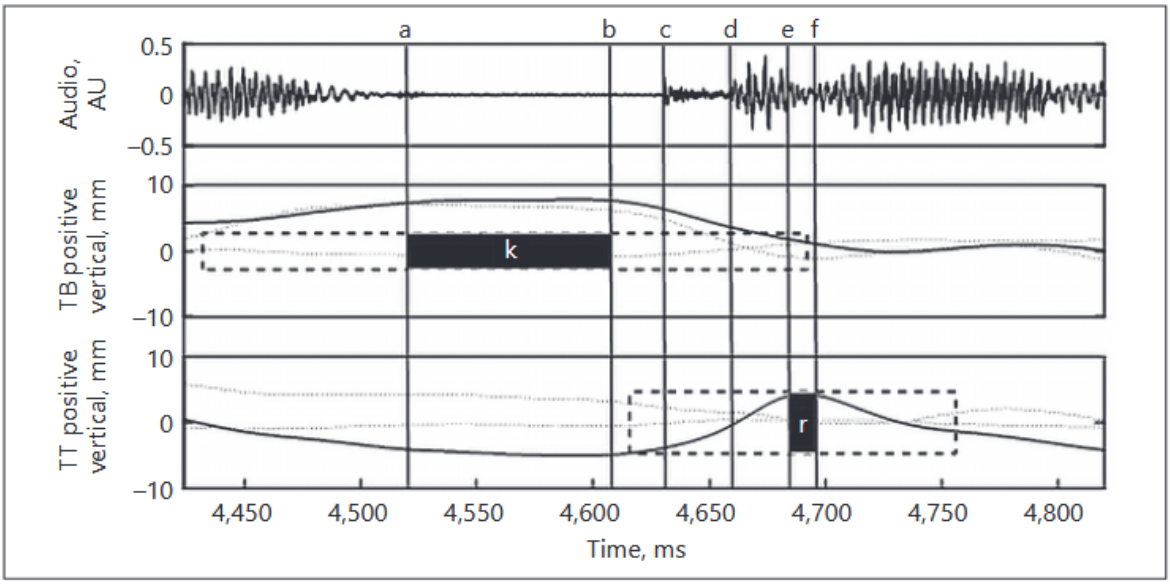

Figure 1.3.2 *Example output from EMA study with Spanish word crema as the token*

TT = tongue tip; TD = tongue dorsum. Reproduced from Gibson et al., 2019.

Chapter 2
The Present Study

This chapter serves two purposes. First, I summarize key findings regarding articulatory overlap in English and Spanish. Using this information, I then make testable predictions about segmentation performance between and within languages.

2.1 Articulatory Overlap in English and Spanish

In this section, I summarize findings from articulatory Phonetics in both languages across three syllabic structures.

2.1.1 Complex Onsets

Complex onsets, defined as sequences of two or more consonants preceding a nuclear vowel, place significant demands on the tongue (Byrd, 1996). Prior research has shown that these coarticulatory demands are handled differently depending on language (e.g., Bombien & Hoole, 2013) and cluster composition (see Pouplier, 2022 for a multi-language analysis). Here, I focus on English and Spanish.

Byrd (1996) offers the following summary of intra-cluster timing in English: "… the closure for the first consonant in a cluster generally is not released until after the closure for the second is formed" (p. 210). This phenomenon has been termed "close transition" (Gibson et al. 2018, p. 102, but see Catford, 1985 for more on this topic), emphasizing the cluster's tight coproduction. To illustrate, I borrow an example from Catford (1985): the words plight and polite are near-homonyms distinguished primarily by the close articulatory transition between /p/ and /l/ in plight (p. 336).

In contrast, Spanish follows a pattern of "open transition" (Gibson et al., 2018, p.102). Gibson et al. (2019) summarize their findings thus: "…Spanish clusters exhibit C1C2 timing patterns of the sort that characterize languages which do not admit clusters as syllable onsets (in that there is a robust interplateau lag between the two consonants in the cluster) …" (p. 2019). Simply put, this means that the first phone of the cluster is released before the second one is formed. Other investigations have reported the same result, including Gibson et al. (2015) and Gibson et al. (2018) (see Crouch et al., 2023 for a report of an analogous phenomenon in Georgian). As it relates to this investigation, the important takeaway is that complex onsets are significantly more overlapped in English than in Spanish.

Moreover, cluster composition appears to influence gestural overlap. I mention three variables here: 1) plosive voicing; 2) presence of fricatives; 3) r-clusters vs. l-clusters. First, voiced plosives in complex onsets (e.g., /bl/, /gr/) tend to be more overlapped than clusters starting with unvoiced stops (e.g., /kl/, /kr/), a finding corroborated in both languages (English: Pouplier et al., 2022; Spanish: Gibson et al., 2019). Conversely, fricative-initial clusters have been shown to be less overlapped, at least in English (Byrd, 1996; Pouplier et al., 2022). Finally,

Gibson et al. (2019) found that l-clusters in Spanish (e.g., /bl/, /kl/) tend to be more overlapped than r-clusters (e.g., /br/, /kr/); notably, this pattern has not been reported in English (Gibson et al., 2018).

2.1.2 Vowel-initial Segments

In this study, VC fragments are of interest since there are different degrees of coarticulation depending on language and coda type.

Prior research has found greater overlap in VC utterances with sonorant codas than those with obstruent codas. Crucially, this pattern has been observed in English but not Spanish. Pouplier et al. (2024) found that vowels followed by nasals have "a significantly and substantially earlier onset of coarticulation in English" (p. 14). These findings comport with Solé's earlier analyses (1992, 1995), showing that English nasalizes vowels completely. Conversely, the same investigations found that Spanish vowels are significantly less nasalized because the lowering of the velum is delayed until the final stages of the vowel (Solé, 1992). For liquid codas, Marin and Pouplier (2010) found "quite large shifts toward the [nuclear] vowel" (p. 392). They attribute this phenomenon to English's dark /l/, which requires an "extreme and earlier vocalic gesture (the tongue body retraction)" (p. 399). No such phenomenon has been reported in Spanish, which has a clear /l/ (Marin & Pouplier, 2014, p. 26).

2.1.3 Simple Onsets

I have found only two published studies on gestural overlap in word-initial simple onsets (CV_ syllables). First, Loevenbruck et al. (1999) studied the timing of gestures across three syllabic structures using English stimuli: kəl, kl, and skl. Results showed significantly less overlap in kəl when compared against the other two, suggesting less coarticulation between consonant (/k/) and vowel (/ə/) versus consonant (/k/) and consonant (/l/). Gu (2025) continued this thread using English and Chinese stimuli. They manipulated consonants in CV stimuli to discern the degree to which changes in sonority predict gestural overlap (e.g., mac vs. back).[2] They found that, holding the vowel constant, more sonorous initial sonorants were associated with shorter temporal gaps (i.e., more overlap) in both languages, a pattern also reported by Crouch et al. (2023) in Georgian.

As it pertains to the present investigation, the important point is that there is no empirical evidence suggesting that gestural overlap in simple onsets differs between English and Spanish because of *language-specific* coarticulatory patterns.[3] However, there is ample evidence that overlap in complex onsets and between vowels and sonorants systematically differ between the two languages.

[2] It is important to note this study assigned sonority values using a continuous scale (borrowed from Parker, 2002) instead of the [± sonorant] binary feature.

[3] However, there is evidence of cross-linguistic similarities in the characteristics of consonants in CV syllables that influence degree of coarticulation (see Recasens, 2015 for a review).

2.2 Outcomes of Interest

In this study, I operationalize PA through a segmentation task where students are asked to segment all the sounds they hear in a spoken word. Responses receive two kinds of scores: 1) initial sound isolated (0/1); 2) unitized response (0/1). As its name suggests, the former denotes a correctly segmented first sound, allowing for regional differences and reasonable substitutions. An example of regional variation would be saying that /a:/ is the first sound in ice for a person speaking Southern American English. Meanwhile, reasonable substitutions are defined as instances where phonetic realization justifies the response (e.g., saying /b/ for the first sound in the Spanish syllable *po*, since Spanish /p/ is perceptually similar to English /b/). Since students had to repeat all stimuli correctly before being prompted to segment them (see Methods in Chapter 4), misheard or mispronounced words are accounted for.

Unitized responses—a term used interchangeably with glued responses—is the other outcome of interest. These are responses where students produce the first two segments of the stimulus without separation. The following are examples: /fæ/ for fan; /kl/ for clasp; /æk/ for act; /ɛl/ for elf. [4] For stimuli with at least three phonemes, repeating the entire word (/fæn/) or including vowels as part of complex onsets (/klæ/ for clasp) is *not* considered a glued response. On the other hand, for items with two phonemes (e.g., eke), repeating the stimulus after a prompt to segment *is* considered a unitized response.

Using both outcomes, I submit the following overarching hypothesis: if coarticulation is the locus of segmentation difficulty, then onsets with more gestural overlap should have lower probabilities of correctly identified initial sounds *and* higher probabilities of unitized responses. In fact, the latter outcome is of special interest. After all, a child might mis-segment the first sound in various ways not obviously related to gestural overlap (e.g., dropping the initial sound). A glued response would be compelling evidence of a coarticulatory effect.

2.3 Research Questions and Hypotheses

Two sets of research questions and hypotheses are detailed below, primary (between-language) and secondary (within-language).

Primary research questions explore between-language differences in three syllabic structures:

> **Research Question 1** (RQ1): Do complex onsets differ in difficulty between English and Spanish?

[4] Responses to items starting with the complex segment /dʒ/ (i.e., jet, drum) are scored as 1 for *initial sound* if the utterance is /dʒ/ or /d/. Responses are scored as 1 for *unitized response* if the utterance is /dʒɛ/ for jet and /dʒr/ or /dr/ for drum. I discuss this further in section 5.4.1.

Hypothesis 1 (H1): Complex onsets in English, relative to Spanish, will be associated with lower probabilities of initial sound segmentation *and* higher probabilities of unitized responses.

Research Question 2 (RQ2): Do vowel-initial stimuli with sonorant codas differ in difficulty between English and Spanish?

Hypothesis 2 (H2): Vowel-initial stimuli with sonorant codas in English will be associated with lower probabilities of initial sound segmentation *and* higher probabilities of glued responses relative to their counterparts in Spanish.

Research Question 3 (RQ3): Do simple onsets differ in difficulty between English and Spanish?

Hypothesis 3 (H3): Simple onsets in both languages will be associated with similar probabilities of initial sound segmentation and glued responses, respectively.

Secondary research questions explore within-language differences in complex onsets:

Research Question 4 (RQ4): Within each language, do voiced and unvoiced plosives differ in difficulty?

Hypothesis 4 (H4): In both languages, complex onsets starting with voiced plosives will have lower probabilities of initial sound segmentation *and* higher probabilities of glued responses.

Research Question 5 (RQ5): Within each language, how do fricative-initial clusters compare in difficulty to plosive-initial clusters?

Hypothesis 5 (H5): In both languages, complex onsets starting with fricatives will have higher probabilities of initial sound segmentation *and* lower probabilities of glued responses.

Research Question 6 (RQ6): Within each language, do l-clusters and r-clusters differ in difficulty?

Hypothesis 6 (H6): In Spanish, l-clusters will be more difficult than r-clusters; this means they will be associated with lower probabilities of initial sound segmentation *and* higher probabilities of glued responses. In English, l-clusters and r-clusters will not differ in difficulty, which means both will be associated with similar probabilities of initial sound segmentation and glued responses, respectively.

Chapter 3
Methods

3.1 Participants

Students were recruited from schools in three US states (Alabama, California, and Wisconsin), as well as the countries of Nicaragua and El Salvador. The seven schools (5 in the United States; 2 in Latin America) employed one of three models which determined the amount of instruction in each language: 100% English; 90% English/10% Spanish; >= 50% Spanish. A total of 474 students participated in the study, spanning TK - Grade 2.

The research protocol for this study was approved by Berkeley's institutional review board. All TK-2 students at participating schools were invited to join the study; however, only those whose parents completed permission forms were offered the opportunity to participate. During the initial assessment session, students were asked if they wanted to participate in the study. None declined.

3.2 Stimuli

This analysis includes responses to a total of 105 monosyllabic stimuli (English: 52, see Appendix A) across a range of onsets (see Variables section). Eight English stimuli were not included in the final analysis because the item bank lacked comparable Spanish counterparts: s-cluster (scrap; scuff; spent; stone; straight), voiceless glottal fricative (hitch); voiced postalveolar affricate (jet); voiced labio-velar approximant (weak). All but one onset type have at least two stimuli per language (see Table 3.2.1). Furthermore, it is noted that Spanish items include pseudowords due to phonotactic constraints, whereas English has no pseudowords.

Table 3.2.1 *Item counts by onset type and language*

Onset Type	English	Spanish
r-Cluster + unvoiced plosive	6	4
r-Cluster + voiced plosive	4	5
l-Cluster + unvoiced plosive	4	5
l-Cluster + voiced plosive	1	4
r-Cluster + fricative	2	2
l-Cluster + fricative	2	2
Vowel-initial + obstruent coda	8	5
Vowel-initial + sonorant coda	4	3
Singleton unvoiced plosive	5	6
Singleton voiced plosive	3	6
Singleton fricative	5	2
Singleton sonorant	8	9
Total	**52**	**53**

3.3 Data Collection

Students completed up to three tasks, the first two of which are not part of the current study. Task order remained the same, and each session lasted about 5 minutes (about 10 for students completing additional measures). I administered all tasks. Whenever possible, students attending a school with literacy instruction in both languages *or* those who reported speaking Spanish at home completed measures in both languages in two separate testing sessions.

Each segmentation task started with 4 trial items, none of which came from the pool of stimuli (see Appendix B). During this time, I said the token, students repeated it, and then I modeled correct segmentation of the stimulus (e.g., clip would be /k/ /l/ /ɪ/ /p/). As part of the measure, students had to repeat the token correctly before being prompted to say all the individual sounds they heard. If students pronounced the stimulus incorrectly, the syllable was repeated until the child said it correctly.

Students were allowed to ask for an item to be repeated up to two times (for a total of three exposures). There was no stop rule and no time limit, but participants could decide to stop at any point.

Items were randomly administered from each language's respective pool. Up to 29 items per language were offered to each participant, though some students were offered as few as 10 due to external constraints (e.g., timing restrictions imposed by the school).

3.4 Score Processing

During data collection at the first two schools, all student responses were recorded. I soon realized that the quality of these recordings did not always allow me to transcribe responses with a high degree of confidence; hence, for the remaining five schools, I decided to transcribe responses in the moment, allowing me to ask students to repeat responses as necessary.

These raw transcriptions were then scored using the following multi-step process. For each item, all unique transcriptions were identified. These unique responses were then compared against a preliminary key and assigned a temporary score (0/1) using custom-built functions in R (R Core Team, 2024). I then manually checked this spreadsheet to ensure that each unique transcription was scored correctly. Having completed this check, this file became the final key against which all individual responses were scored. This process was repeated for each language/outcome combination.

3.5 Variables

Initial Sound is a binary variable where 1 denotes the participant has correctly identified and isolated the item's first segment.

Unitized Response is a binary variable where 1 denotes the participant has produced the first two phones of the stimulus without separation.

Onset Type is a categorical variable with 12 levels based on onset structure: r-cluster + unvoiced plosive; r-cluster + voiced plosive; l-cluster + unvoiced plosive; l-cluster + voiced plosive; r-cluster + fricative; l-cluster + fricative; VC obstruent (vowel-initial + obstruent coda); VC sonorant (vowel-initial + sonorant coda); singleton unvoiced plosive; singleton voiced plosive; singleton fricative; singleton sonorant.

English is a dummy variable denoting whether the stimulus was in English. The reference group is Spanish.

Coda Cluster is a dummy variable denoting whether the stimulus has a coda with more than 1 sound. The reference group is a singleton coda.

Vowel Type is a categorical variable with 4 levels: high, middle, low, and diphthong. The reference group is high.

Real Word is a dummy variable denoting whether the stimulus is a real word. The reference group is no (i.e., the stimulus is not a real word).

Phonemes is a continuous variable denoting the number of phonemes in the stimulus, centered at 2.

Log Frequency is a continuous, standardized variable denoting the log-transformed frequency of the stimulus. English values are from Balota et al. (2007); Spanish values are from Duñabeita et al. (2010). The variable was standardized within language.

Grade is a categorical variable denoting the participant's grade level: K, 1, 2, or TK. The reference group is K. Since it was not possible to collect information about the children's date of birth, this variable serves as a proxy for age.

Spanish at Home is a dummy variable denoting whether the participant reported speaking Spanish at home. The reference group is Spanish not spoken at home.

Program Type is a categorical variable with 3 levels denoting the percent of instruction in each language in a given school-day: 100% English; 90% English/10% Spanish; >= 50% Spanish. The reference group is 100% English.

3.6 Analysis

A Generalized Linear Mixed Model (GLMM, Rabe-Hesketh & Skrondal, 2022) framework was employed to answer all research questions. This approach allows for a complex random effect structure due to nested data, including crossed random effects. Moreover, item-side and person-side predictors can be included to explain relevant variance. In this analysis, effects of item characteristics are of primary interest, while person attributes are control variables. It is noted that these models are parametrically equivalent to Explanatory Item Response Theory models (EIRT; for more on the topic, see De Boeck & Wilson, 2004).

3.6.1 Model Building

The models described below were fitted for each outcome of interest.

I started by fitting a three-level random-intercept logistic regression model (Model 1):

$$\text{logit}\,(P(Y_{ijk} = 1 \mid \zeta_k, \theta_{1j}, \theta_{2j}, ...)) = \beta_0 + \text{English}_i \times \theta_{1j}^{(2)} + \text{Spanish}_i \times \theta_{2j}^{(2)} + \zeta_i^{(2)} + \zeta_k^{(3)} \quad \textbf{Equation 1}$$

where *i* denotes item, *j* denotes person, and *k* denotes classroom. English_i and Spanish_i are dummy variables denoting the language of the stimuli. The corresponding latent abilities in each language, $\theta_{1j}^{(2)}$ and $\theta_{2j}^{(2)}$, are assumed to have a bivariate normal distribution with means of 0 and variances $\psi_{11}^{(2)}$ and $\psi_{22}^{(2)}$, respectively, and covariance $\psi_{21}^{(2)}$. The random item and classroom effects, $\zeta_i^{(2)}$ and $\zeta_k^{(3)}$, are assumed to be normally distributed with means of 0 and variances $\psi_3^{(2)}$ and $\psi_4^{(3)}$, respectively.

A saturated model on the item side (Model 2) was then fitted to estimate an easiness parameter for each item:

$$\text{logit}\,(P(Y_{ijk} = 1 \mid \zeta_k, \theta_{1j}, \theta_{2j}, ...)) = \beta_1\textbf{Item}_i + \text{English}_i \times \theta_{1j}^{(2)} + \text{Spanish}_i \times \theta_{2j}^{(2)} + \zeta_k^{(3)} \quad \textbf{Equation 2}$$

where Item_i is a vector of dummy variables and β_1 are corresponding regression coefficients representing item easiness. Note that there is no reference group in order to estimate a parameter for each item. All other parameters remain as defined in Equation 1.

Subsequently, a model with all item and person covariates (Model 3) was fitted that also includes an interaction between **Onset Type** and English:

$$\text{logit}\,(P(Y_{ijk} = 1 \mid \zeta_k, \theta_{1j}, \theta_{2j}, ...)) = \beta_1\textbf{OnsetType}_i + \beta_2\textbf{OnsetType}_i \times \text{English}_i + \beta_3\textbf{ItemPredictors}_i + \beta_4\textbf{PersonPredictors}_j + \text{English}_i \times \theta_{1j}^{(2)} + \text{Spanish}_i \times \theta_{2j}^{(2)} + \zeta_i^{(2)} + \zeta_k^{(3)} \quad \textbf{Equation 3}$$

where **Onset Type** is a vector of dummy variables capturing all 12 onset types of interest; $\textbf{ItemPredictors}_i$ is a vector of the remaining item-side variables; and $\textbf{PersonPredictors}_j$ is a

vector of all person-side variables. It is noted that **Onset Type** does not have a reference group in this parametrization. All other variables remain as defined in Equations 1 and 2. Since crossed random effects models are difficult to estimate, these models were estimated using the Laplace approximation as employed by the lme4 package (Bates et al., 2015).

The final models—Model 4-Initial and Model 4-Unitized—dropped the random item effect because several indices suggested it was not necessary (as described below). These models were fitted in Stata[5] (StataCorp, 2023) since adaptive quadrature is more accurate than the Laplace approximation (Skrondal & Rabe-Hesketh, 2004, Chapter 6). They were parametrized as follows:

$$\text{logit}\left(P(Y_{ijk}=1 \mid \zeta_k, \theta_{1j}, \theta_{2j}, \ldots)\right) = \beta_1\mathbf{OnsetType}_i + \beta_2\mathbf{OnsetType}_i \times \text{English}_i + \beta_3\mathbf{ItemPredictors}_i + \beta_4\mathbf{PersonPredictors}_j + \text{English}_i \times \theta_{1j}^{(2)} + \text{Spanish}_i \times \theta_{2j}^{(2)} + \zeta_k^{(3)} \quad \textbf{Equation 4}$$

with all variables as previously defined.

The three indicators used to justify dropping the random item effect were:

1) The proportion of between-item variance explained by Model 3, defined as: ($\psi_3^{(2)}$ from Equation 1 – $\psi_3^{(2)}$ from Equation 3) / $\psi_3^{(2)}$ from Equation 1. A higher proportion of explained variance indicates greater explanatory power of the item covariates.

2) The correlation between item easiness parameters from Model 4 and those from Model 2. Item easiness was reconstructed using a linear combination of Model 4's covariate parameters (weighted by the corresponding coefficient estimates) and then correlated with item parameters from Model 2. A higher correlation indicates that item predictors explain item easiness.

3) The Root Mean Square Difference (RMSD) between Model 4's reconstructed item easiness (as described above) and Model 2's parameters. A smaller RMSD (in logits) indicates better parameter recovery.

3.6.2 Parameters of Interest

Across both models, parameters represent the log-odds (and log-odds differences associated with covariates) of producing that response type conditioning on all variables and random effects. In Model 4-Initial, higher β_1 values denote easier items. The opposite is true in Model 4-Unitized, where higher values denote more difficult stimuli (i.e., those more likely to yield a glued response). The level of significance is set to .05.

[5] These models were fitted with 7 integration points. When I attempted to increase this number, the models ran too slowly to be estimated.

Five out of the six research questions are addressed by asking whether an average of estimated parameters significantly differs from zero (see below for details). These averages were estimated using Stata's *lincom* command (StataCorp, 2023), a function which also estimates corresponding z statistic. These estimates are reported in Chapter 4. Finally, to facilitate interpretation and discussion in Chapter 5, marginal probabilities were estimated using Stata's *margins* command (StataCorp, 2023).

The first research question is addressed by averaging 6 of the interaction parameters in β_2 (see Equation 4). This result represents the mean difference between English and Spanish clusters of the same type. If English clusters are more difficult, the difference will be statistically significantly less than zero for first sound isolation and greater than zero for unitized response.

The second research question is answered directly by the one interaction parameter in β_2 for *VC Sonorant*. This interaction term represents the difference between English and Spanish word-initial stimuli with sonorant codas. If English counterparts are more difficult on average, the parameter will be statistically significantly less than zero for first sound isolation and greater than zero for unitized response.

The third research question is addressed by averaging 4 interaction parameters in β_2 for the singleton onsets. This result represents the mean difference of English singletons compared to Spanish. If singletons are equally difficult on average in both languages, the difference will be indistinguishable from zero in both outcomes.

The fourth research question is addressed by subtracting the average of unvoiced plosive clusters (2) from the average of voiced plosive clusters (2) for each language. This result represents the mean difference between voiced plosive and unvoiced plosive clusters within a language. If voiced plosives are more difficult on average, the difference will be statistically significantly less than zero for first sound isolation and greater than zero for unitized response.

The fifth research question is addressed by subtracting the average of plosive-initial clusters (4) from the average of fricative clusters (2) for each language. This result represents the mean difference between fricative and plosive clusters within each language. If fricative clusters are easier, the difference will be statistically significantly greater than zero for first sound isolation and less than zero for unitized response.

The sixth research question is addressed by subtracting the average of r-clusters (3) from the average of l-clusters (3) for each language. This result represents the mean difference between l- and r-clusters within each language. If l-clusters are more difficult on average, the difference will be statistically significantly less than zero for first sound isolation and greater than zero for unitized response.

The code used to estimate all comparisons can be found in Appendix C.

Chapter 4
Methods

4.1 Descriptive Statistics

4.1.1 Items

The analytical sample is composed of 105 items (English: 52). Tables 4.1.1.1 and 4.1.1.2 show the descriptive statistics and counts for item-side control variables. As described in Chapter 3, the variable *Log Frequency* was constructed by standardizing the log-transformed frequency of the stimulus within each language. English values are from Balota et al. (2007); Spanish values are from Duñabeita et al. (2010).

Table 4.1.1.1 *Descriptive statistics for item-side control variables*

	M	***SD***	**Min**	**Max**
Phonemes	1.01	.93	0	3
Log Frequency	0.00	.99	-3.12	2.16

Table 4.1.1.2 *Counts for item-side control variables*

	N (%)
English	52 (49.5%)
Vowel Type	
High	22 (21%)
Middle	56 (53%)
Low	12 (11%)
Diphthong	15 (14%)
Coda Cluster	17 (16%)
Real Word	58 (55%)

4.1.2 Persons

The analytical sample is composed of 474 students from 72 classrooms (*M* = 6.6, *SD* = 3.6, range: 2-16). Table 4.1.2.1 shows the counts for person-side control variables. On average, participants responded to 26 items in both languages. Across 12,388 responses, the accuracy rate of first sound isolation was 83% (*M* = .83, *SD* = .38), while unitized responses occurred in 10% of cases (*M* = .10, *SD* = .30).

Table 4.1.2.1 *Counts for person-side control variables*

	N (%)
Students	474 (100%)
Grade	
Kindergarten	104 (22%)
Grade 1	176 (37%)
Grade 2	179 (38%)
TK	15 (3%)
Program Type	
100% English	175 (37%)
90% English	205 (43%)
>=50% Spanish	94 (20%)
Spanish spoken at home	293 (62%)

4.2 Item Parameter Recovery Indices

As can be seen from Table 4.2.1, the item-side predictors from Model 3 explain about 90% of the between-item variance in both outcomes. Collapsing across languages, the correlation of predicted item difficulties from Models 2 and 3 is .87 and .90 for first sound and unitized sound, respectively (r columns in Table 4.2.2). RMSD estimates range from .25 to .48 logits (RMSD columns in Table 4.2.2). Altogether, I interpret these results as evidence that item-side predictors are indeed explaining between-item differences in a meaningful way.

Table 4.2.1 *Estimates of item variance (Model 1) and residual variance (Model 3) (both in logits) and variance explained (%)*

	Model 1	**Model 3**	**Percent Explained**
First Sound ($\psi_3^{(2)}$)	1.13	.08	93
Unitized Sound ($\psi_3^{(2)}$)	1.74	.18	90

Table 4.2.2 *Correlation (r) and RMSD (in logits) of item parameters from Model 2 and Model 3*

	Overall		**English**		**Spanish**	
	r	RMSD	r	RMSD	r	RMSD
First Sound	.90	.25	.91	.22	.89	.29
Unitized Sound	.87	.48	.92	.40	.79	.55

4.3 Primary Research Questions

Table 4.3.1 summarizes results for the primary research questions. Appendix D lists the full output for Model 4-Initial (Column 1) and Model 4-Unitized (Column 2).

The first research question asked whether English onset clusters would be more difficult to segment than Spanish ones due to higher degrees of coarticulation in English clusters compared to Spanish. Results indicate that the log-odds of correctly isolating the first sound of an English cluster, averaging across all six cluster types, is estimated to be .61 logits ($p < .05$) lower than in Spanish, holding all other predictors constant. Crucially, these English onsets also appear to be more likely to elicit a glued response than their Spanish counterparts by an estimated .94 logits ($p < .01$).

It is important to note that the estimates reported above, by averaging across all six cluster types, might overgeneralize differences that are cluster specific. An examination of point estimates from Appendix D suggests that, for first sound segmentation, voiced plosives and fricative-initial clusters in particular elicited disproportionately more errors in English (vs. Spanish). Similarly, it appears that all but two English complex onsets are more likely to elicit unitized responses: l-clusters with initial consonants that were either unvoiced plosives or fricatives.

The second research investigated asked whether vowels followed by sonorants would be more difficult to segment in English than Spanish due to the increased gestural overlap of the former. Results show that correctly isolating the vowel in English stimuli with sonorant codas is estimated to be almost 1.8 logits more difficult than in Spanish ($p < .001$). These same items are also about 1.11 logits more likely to produce unitized responses in English ($p < .05$).

Here, it is worth unpacking the point estimates for vowel-initial stimuli with sonorant and obstruent codas between languages. While first sound isolation also appears to be more difficult for English items with obstruent codas relative to Spanish ($\hat{\beta} = -.79$, $p < .05$), the difference between English obstruent codas and English sonorant codas is substantial: $\hat{\beta} = 1.01$ ($p = .001$).[6] In other words, the difference in first sound segmentation gap between English and Spanish is greater for sonorant codas than for obstruent codas. Second, and of particular interest for this study, glued responses are statistically indistinguishable for vowel-initial stimuli with obstruent codas between languages: $\hat{\beta} = -.40$ ($p = .35$). Thus, while first sound isolation appears to be more difficult across all vowel-initial syllables in English (vs. Spanish), the sonorant coda subtype in particular tends to be more difficult than its within-language (English vowel-initial stimuli with obstruent codas) and between-language (Spanish vowel-initial stimuli with sonorant codas) counterparts.

The third research question tested if simple onsets differ between languages since there is no evidence of language-specific differences in coarticulation between English and Spanish CV syllables. Results found no evidence that English singletons, averaged across the four classes of

[6] This is a post-hoc, unplanned comparison of English x *VC Obstruent* ($\hat{\beta} = -.79$, $p = .02$) and English x *VC Sonorant* ($\hat{\beta} = -1.80$, $p < .001$) (see Appendix D).

initial phonemes, differ from Spanish with respect to initial sound isolation (.01, p = .98) or unitized responses (-.38, p = .27).

Table 4.3.1 *Estimates (in logits) related to primary research questions*

	RQ1		**RQ2**		**RQ3**	
	Est (*SE*)	*p*	Est (*SE*)	*p*	Est (*SE*)	*p*
First Sound	**-.61** (.26)	.02	**-1.80** (.37)	< .001	.01 (.29)	.98
Unitized Sound	**.94** (.30)	< .01	**1.11** (.48)	.02	-.38 (.35)	.27

4.4 Secondary Research Questions

Table 4.4.1 summarizes results for secondary research questions, all of which are within-language.

The fourth research question addressed whether the greater gestural overlap in voiced-plosive clusters makes them harder to segment than unvoiced-plosive onsets. In English, the mean of first sound isolation is about 1.09 logits lower for voiced versus unvoiced stops (p < .001); in Spanish, the estimated difference is .67 logits (p < .001). These same clusters are also significantly more likely to elicit glued responses in both languages (English: 1.06, p < .001; Spanish: .39, p = .02).

The fifth research question investigated whether fricative-initial clusters are easier to segment than plosive-initial ones, given that the former exhibit less gestural overlap. In English, the average difference in first sound isolation between fricative and plosive clusters is estimated as .84 logits (p < .001); in Spanish, the estimated difference is greater still, about 1.27 logits (p < .001). Unitized responses are also less likely for fricative clusters in both languages (compared to within-language, plosive clusters): English: -1.33 (p < .001); Spanish: -1.12 (p < .001). Taken together, these results suggest that fricative clusters are easier in both languages.

The sixth and final research question asked whether l-clusters differ from r-clusters since there is evidence that l-clusters exhibit more overlap in Spanish but not necessarily in English. Results found no statistically significant evidence that l-clusters differ from r-clusters in either language (see RQ6 in Table 4.4.1). It is important to note that this planned contrast compared the average of all l-clusters to the average of all r-clusters, both of which include fricative-initial stimuli. Given the large effect of fricative clusters reported above, it stands to reason that comparisons including such items may obscure underlying differences. A post-hoc comparison excluding fricative clusters found no such difference (see RQ6b in Table 4.4.1).

Table 4.4.1 *Estimates (in logits) related to secondary research questions*

	English		**Spanish**	
	Est (*SE*)	*p*	Est (*SE*)	*p*
RQ4				
First Sound	**-1.09** (.17)	< .001	**-.67** (.15)	< .001
Unitized Sound	**1.06** (.19)	< .001	**.39** (.17)	.02
RQ5				
First Sound	**.84** (.15)	< .001	**1.27** (.18)	< .001
Unitized Sound	**-1.36** (.17)	< .001	**-1.12** (.20)	< .001
RQ6				
First Sound	.16 (.15)	.25	-.03 (.14)	.81
Unitized Sound	-.29 (.17)	.08	.20 (.16)	.21
RQ6b				
First Sound	-.03 (.16)	.85	-.10 (.14)	.46
Unitized Sound	-.28 (.18)	.11	.22 (.16)	.17

Chapter 5
Discussion

This investigation set out to explore whether gestural overlap, as a proxy for coarticulation, could predict segmentation performance in beginning readers. First, I used results from articulatory studies in English and Spanish to identify phonological structures associated with more overlap—both between and within languages. These articulatory patterns informed hypotheses about the kinds of stimuli that would be difficult to separate, leading to an experiment where beginning readers segmented orally presented stimuli. Multilevel logistic regressions were used to ascertain whether response behaviors could be predicted from these phonological structures. Overall, results suggest that gestural overlap predicts segmentation performance.

5.1 The Two Outcomes

Before interpreting specific results, I would like to make several points about the two outcomes that were modeled: initial sound isolation and unitization. From the beginning, these response behaviors were hypothesized to exhibit a negative relationship: as the probability of correct first sound decreases, the likelihood of a unitized response should increase. Importantly, this relationship was far from guaranteed. After all, there are various behaviors that would all indicate failure to identify the first sound (e.g., dropping the first sound, replacing it with an unrelated phoneme, etc.) that are also not the production of an item's first two phonemes.

As explained in Chapter 2, the most compelling results would be ones where rates of glued responses are higher in contexts with higher (vs. lower) degrees of coarticulation, not simply ones where rates of first sound isolation are lower. This latter point is especially important since the rate of failure for first sound isolation (.17 = 1 - .83, from Chapter 4) was almost double that of glued responses (.10) on average. This confirms that unitization was not the only cause of first sound difficulty. A future study, or even a reanalysis of these same data, could explore the kinds and relative frequency of other segmentation errors.

With these points in mind, I now turn to a discussion of each research question and corresponding hypothesis.

5.2 Between Language Comparisons

The first set of research questions explored between-language patterns. Based on prior articulatory studies (Byrd, 1996; Gibson et al., 2019), it was predicted that increased gestural overlap in English complex onsets would make segmentation more difficult than the open transition found in Spanish clusters (the first hypothesis). Results support this hypothesis, as the average English cluster is more likely to elicit a glued response. In terms of marginal probability, this difference translates to about .1 (see Figure 5.2.1). It is compelling that the likelihood of unitized responses is statistically greater for all but two of the interaction terms. Of those two,

one (l-clusters with voiced plosive initials) has a point estimate with a p-value trending toward significance ($\hat{\beta}$ = .60, p < .11) on the strength of a single stimulus (see Appendix D).

The second hypothesis posited that English vowel-initial stimuli with sonorant codas would be more difficult than their Spanish counterparts since English sonorants adhere more closely to the preceding vowel (Marin & Pouplier, 2014; Pouplier et al., 2024) (the second hyptothesis). Results again support this hypothesis. As can be seen in Figure 5.2.2, the marginal probability of producing a glued response for English vowel-initial items with sonorant codas (.066) was more than twice that of Spanish (.028). Crucially, stimuli with obstruent codas in English and Spanish were not more likely to produce unitized responses. This suggests that gestural overlap is the source of difficulty for such stimuli with sonorant codas, manifesting in a higher rate of whole-word responses (e.g., /il/ for eel).

An unexpected finding regarding vowel-initial stimuli was the markedly greater difficulty of English versus Spanish stimuli. Considering the aforementioned result for English vowel-initial stimuli with sonorant codas, the language gap is notable: about .17 (see Figure 5.2.2). But even for counterparts with obstruent codas—where there was no empirical basis to predict that English would be more difficult—Spanish items have a .06 advantage. As described above, this difference cannot be attributed to increased unitization since this interaction term was not statistically significant.

Though results from this experiment cannot explain why vowel-initial stimuli appear to be more problematic in English, one observation is worth making: there was a tendency for students to start their responses to these English items—and not Spanish—with a schwa (/ə/) (e.g., /ə/ /aɪ/ /s/ for ice). It seems plausible that whatever the source of difficulty may be, it also contributed to the frequency of this behavior.

A third prediction was that there would be no between-language differences for stimuli with singleton onsets (the third hypothesis). Unlike the first two research questions which examined specific differences based on articulatory Phonetics, this one tested whether the absence of such evidence could be interpreted as a null difference. Results are consistent with that interpretation. The average interaction terms of English and simplex onset types were not statistically significant for either outcome (see Figure 5.2.3). Examining individual parameter estimates (instead of their average) does not change this conclusion: of the four interaction terms, none reached significance (see Appendix D).

Taken as a whole, these findings are consistent with the idea that greater articulatory overlap is associated with greater segmentation difficulty. Specifically, English clusters and English sonorant codas—both of which have been reported to exhibit more articulatory overlap than their counterparts in Spanish—were associated with higher probabilities of unitization (and lower rates of initial sound isolation).

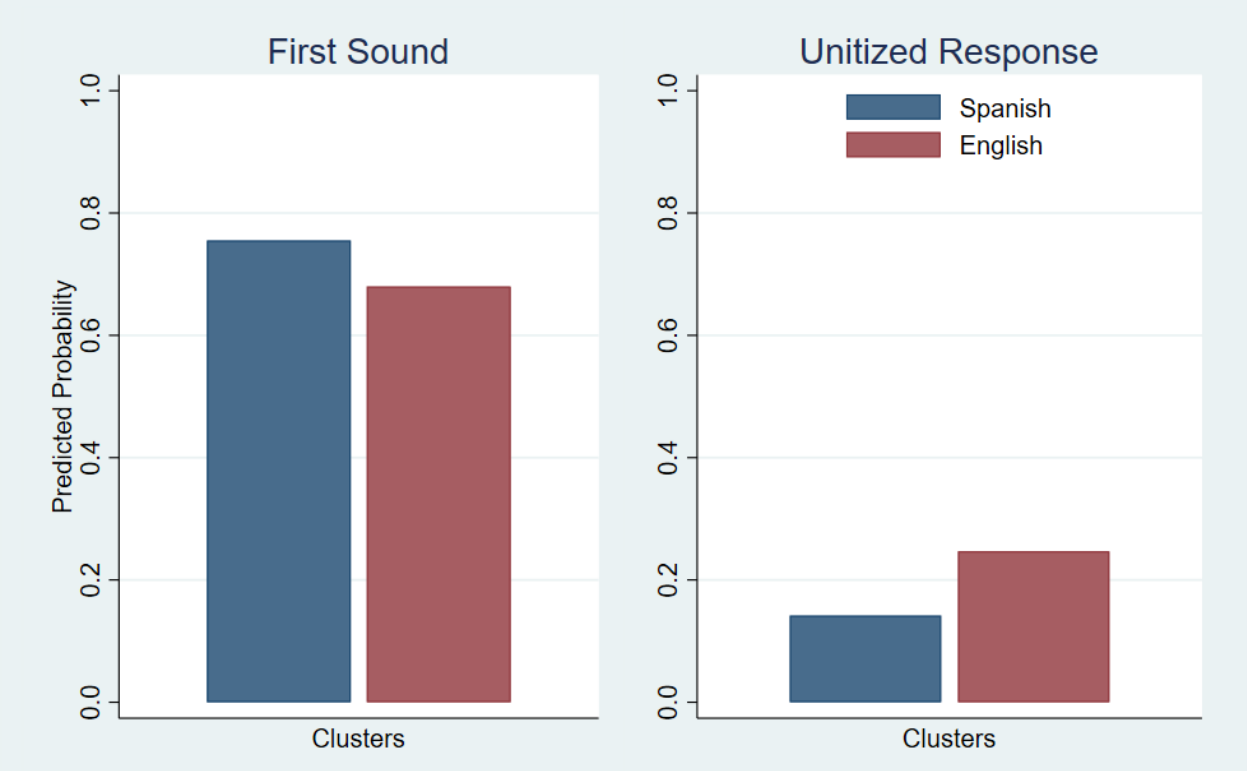

Figure 5.2.1 *Marginal probabilities related to the difficulty of segmenting English and Spanish clusters (RQ1/H1).* For first sound, the marginal probability represents the estimated likelihood of correctly identifying the initial phone of a stimulus (as defined in Chapter 3). For unitized response, the marginal probability represents the estimated likelihood of producing the first two segments of the stimulus (as defined in Chapter 3).

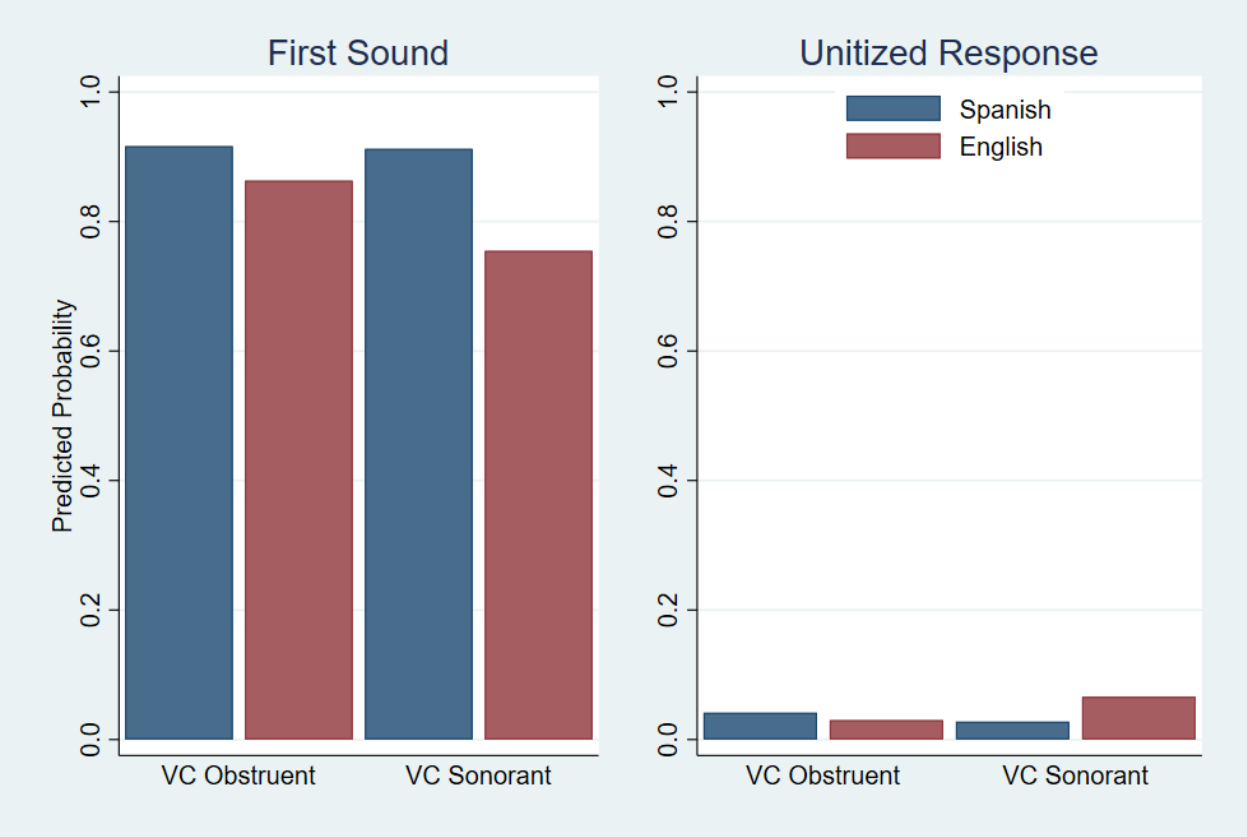

Figure 5.2.2 *Marginal probabilities related to the difficulty of segmenting vowel-initial stimuli with different coda types in English and Spanish (RQ2/H2).* For first sound, the marginal probability represents the estimated likelihood of correctly identifying the initial phone of a stimulus (as defined in Chapter 3). For unitized response, the marginal probability represents the estimated likelihood of producing the first two segments of the stimulus (as defined in Chapter 3).

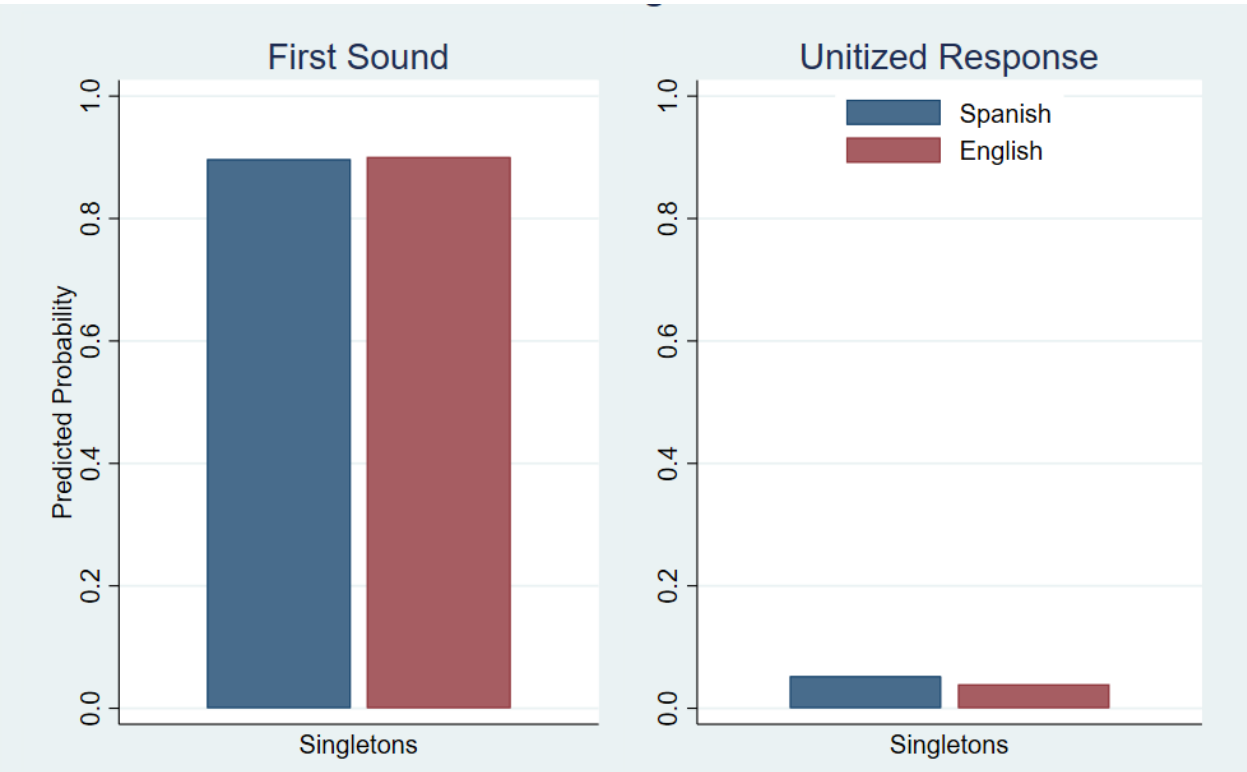

Figure 5.2.3 *Marginal probabilities related to the difficulty of segmenting simplex onsets in English and Spanish (RQ3/H3).* For first sound, the marginal probability represents the estimated likelihood of correctly identifying the initial phone of a stimulus (as defined in Chapter 3). For unitized response, the marginal probability represents the estimated likelihood of producing the first two segments of the stimulus (as defined in Chapter 3).

5.3 Within Language Patterns

The second set of research questions explored within-language patterns. Prior studies have found that voiced plosives tend to be more overlapped with the adjacent liquid (Gibson et al., 2019; Pouplier et al., 2022). Thus, it was expected that they would be more difficult to segment than their unvoiced counterparts in both languages (the fourth hypothesis). As can be seen in Figure 5.3.1, clusters starting with voiced stops were almost twice as likely to elicit glued responses in English (an estimated difference of about .15). The difference was smaller within Spanish (about .05), but still significant as reported in Chapter 4. These findings all support the fourth hypothesis.

Still, the difference in voicing effect between languages is notable. One explanation for the attenuation in Spanish could relate to the phenomenon of svarabhakti vowels (also known as epenthetic vowels)—the addition of vocalic fragments that do not change a word's meaning (Bradley, 2007; Ramírez, 2012). In Spanish clusters, epenthetic vowels have been shown to occur most frequently after voiced stops and after r-clusters (Colantoni and Steele, 2007, 2011; Gibson et al., 2019). It could be that the presence of these vocalic fragments mitigates the coarticulation between members of the cluster, leading to fewer instances of glued responses. In other words: voiced plosives are more overlapped in both languages, but they are less coarticulated than they otherwise would be in Spanish because of the intervening vowel.

Turning now to the difference between plosive and fricative clusters, it was predicted that the latter would be easier in both languages (the fifth hypothesis). As can be seen in Figure 5.3.2, complex onsets starting with /f/ had about half the probability of eliciting unitized responses within each language (in comparison to plosive clusters). Unlike the effect of voicing with plosives, which seemed to be moderated by language, the impact of fricatives in clusters appears to be more consistent between languages. One possibility for this parsimony—and most importantly, for the relative ease of segmenting fricative clusters—is that frication itself might inadvertently attenuate coarticulation with adjacent consonants (Cutting, 1975).

My sixth and final hypothesis was that l-clusters would be more difficult to segment than r-clusters in Spanish but not English (the sixth hypothesis) since Gibson et al. (2019) found l-clusters to be more overlapped in the former. There was no evidence supporting this prediction. Figure 5.3.3, which illustrates the marginal probabilities corresponding to the logit estimates reported in Chapter 4, shows no significant differences between the two cluster types (for either outcome) in either language. One possible explanation for this null finding is that the difference in gestural overlap between l- and r-clusters is simply too small, perhaps due to the presence of the aforementioned svarabhakti vowels.

Finding supporting evidence for two out of three hypotheses suggests that within-language differences in articulatory patterns also influence segmentation behavior.

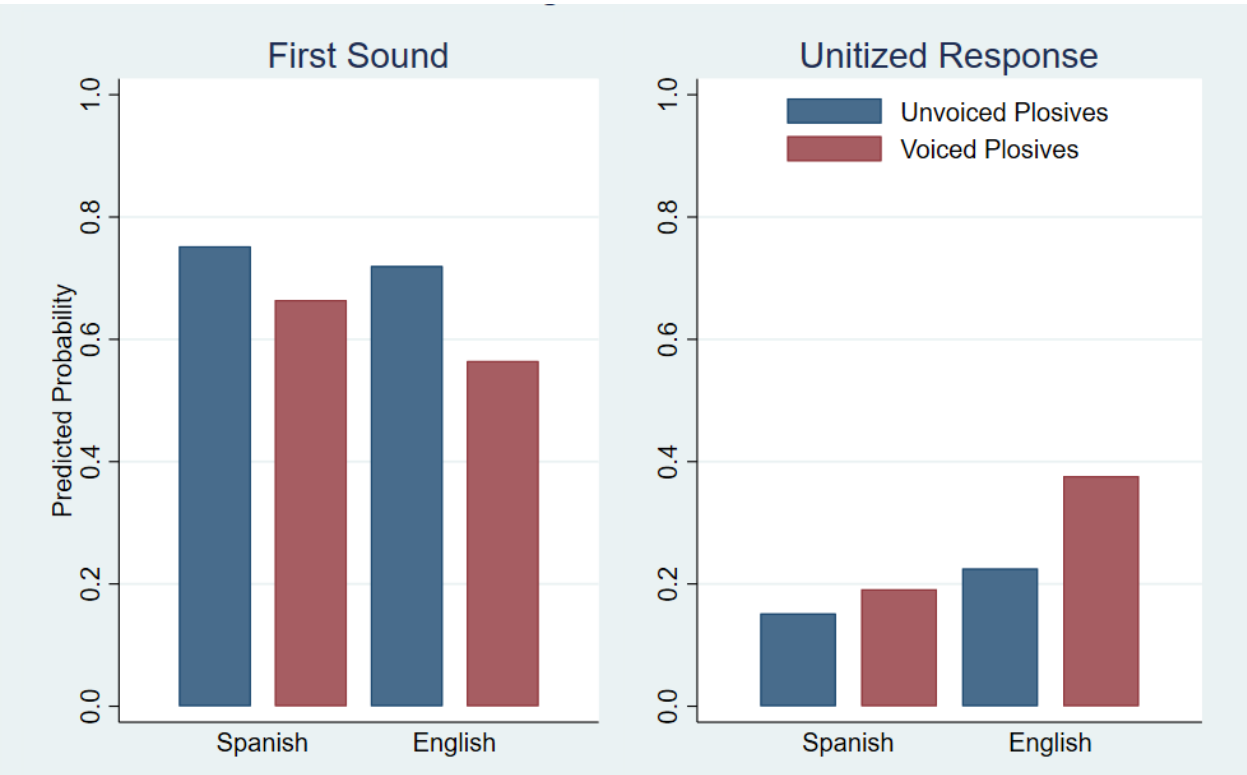

Figure 5.3.1 *Marginal probabilities related to the difficulty of segmenting plosive clusters by voicing type in English and Spanish (RQ4/H4).* For first sound, the marginal probability represents the estimated likelihood of correctly identifying the initial phone of a stimulus (as defined in Chapter 3). For unitized response, the marginal probability represents the estimated likelihood of producing the first two segments of the stimulus (as defined in Chapter 3).

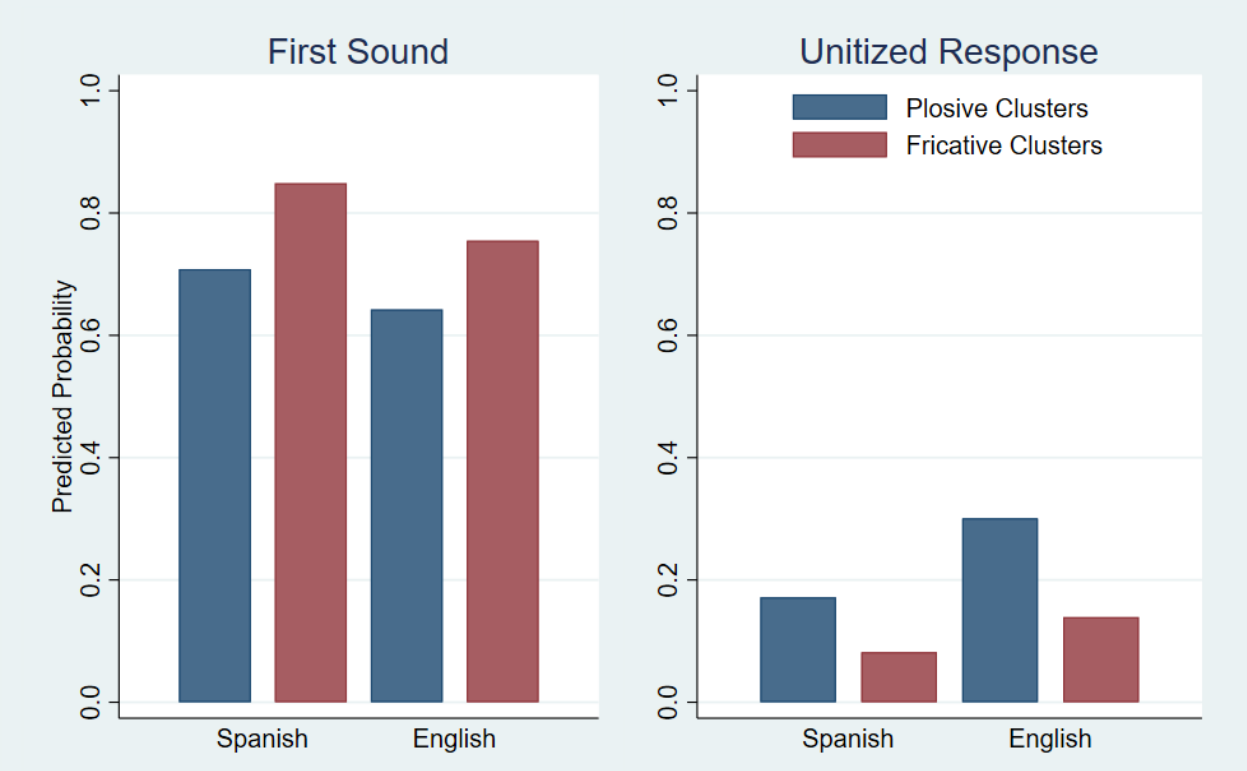

Figure 5.3.2 *Marginal probabilities related to the difficulty of segmenting plosive versus fricative clusters in English and Spanish (RQ5/H5).* For first sound, the marginal probability represents the estimated likelihood of correctly identifying the initial phone of a stimulus (as defined in Chapter 3). For unitized response, the marginal probability represents the estimated likelihood of producing the first two segments of the stimulus (as defined in Chapter 3).

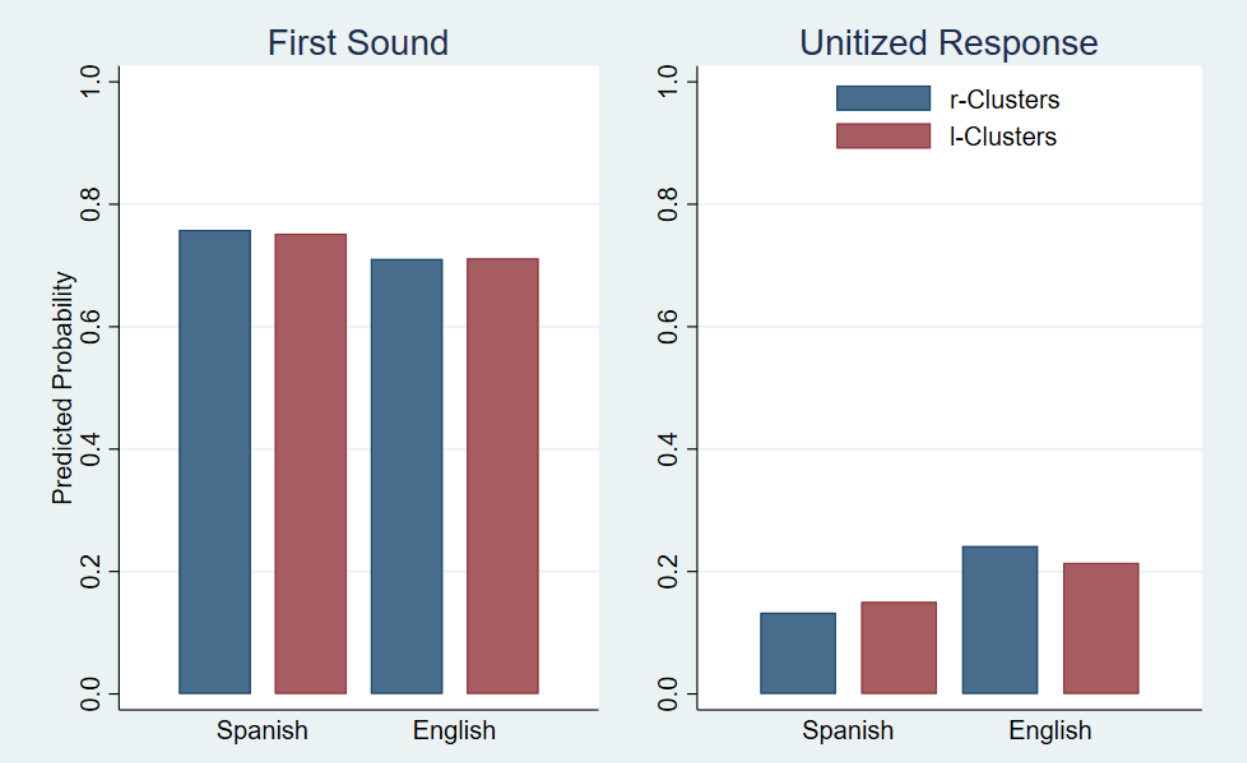

Figure 5.3.3 *Marginal probabilities of segmenting r-clusters versus l-clusters in English and Spanish (RQ6/H6).* For first sound, the marginal probability represents the estimated likelihood of correctly identifying the initial phone of a stimulus (as defined in Chapter 3). For unitized response, the marginal probability represents the estimated likelihood of producing the first two segments of the stimulus (as defined in Chapter 3).

5.4 Other Observations

In this section, I discuss two sets of observations not covered by either set of research questions.

5.4.1 Other Segmentation Behaviors

First, items with diphthongs sometimes elicited what appears to be target-by-target segmentation: instances where the vowel was separated into its individual elements (e.g., /a/ /ɪ/ for ice). This occurred both in English and Spanish, suggesting that at least some children are sensitive to the individual gestures within these fragments. More generally, I interpret this phenomenon as additional evidence that articulation—in this case, the transition from one vocalic segment to another—informs segmentation.

Conversely, one item (ño) with the Spanish palatal nasal /ɲ/ was consistently either segmented as a single sound (/ɲ/, /n/, or /j/) or unitized (/ɲo/). There were no instances of /n/ /j/. Crucially, although Massone (1988) suggested that the phonetic realization of this phoneme can be thought of as a two-gesture complex sequence of /n/ /j/, the literature generally treats this segment as having a single gesture (Ladefoged & Maddieson, 1996; Tomás, 1932). The absence of /n/ /j/ segmentation would comport with the latter interpretation, but the handful of /n/ and /j/ single-phone responses lend some support to the idea that at least some individuals perceive /ɲ/ as having more than 1 articulatory target.

Along similar lines, the voiced post-alveolar affricate in English /dʒ/ elicited similar behavior. Out of 96 scorable responses to the stimulus jet, there was not a single instance where the first sound was segmented as /d/ /ʒ/, and only one where it was identified as /d/. In every other instance, the item was either segmented as /dʒ/ (which was scored as a correct first sound) or /dʒɛ/ (scored as a unitized response). Unlike /ɲ/, /dʒ/ is typically considered a complex segment composed of two distinct gestures (Ladefoged & Maddieson, 1996). One possible explanation for this discrepancy between phonetic reality and segmentation behavior is that the two gestures are so tightly executed that students perceive them as a single unit.

I explored this topic further by examining raw scores for the 96 responses to the stimulus jet. Averages were .95 for first sound and .01 for glued responses. Compared to the sample-wide averages reported in Chapter 4, these means are well above average (for first sound) and well below average (for unitized responses), suggesting a relatively easy onset. In contrast, the affricated dr cluster (in the stimulus drum) had averages of .44 for first sound (/dʒ/ or /d/ since there was no instance of /d/ /ʒ/) and .48 for unitization (/dʒr/ or /dr/) across 101 responses—one of the most difficult stimuli. Considering the above, this would suggest that the difficulty of this cluster type stems from the increased overlap between the affricated initial sound and the adjacent rhotic—not the initial consonant.

The phenomena described above, viewed in isolation, are ostensibly explained by the orthography effect, a phenomenon whereby knowledge of spelling influences speech perception (i.e., Ziegler & Ferrand, 1998). Seen in this light, the target-by-target segmentation of diphthongs and the unitization of /dʒ/ (and possibly /ɲ/) could be attributed to their corresponding graphemic

representations. Diphthongs, in this sample of stimuli, were all spelled with more than one letter (though not necessarily adjacent, e.g., ice); meanwhile, the consonants in question were all spelled with a single letter. Note, however, that this account fails to explain the unitization of clusters, where each consonant was always spelled with its own grapheme.

5.4.2 Onset Age-of-Acquisition

Originally, I had planned to include an additional item-level predictor: onset age-of-acquisition (hereafter, onset AoA). Since stimuli were not matched on this sub-lexical variable, it seemed prudent to control for differences in the normative acquisition of onsets. After all, participants spanned various grades and some onsets might not have been fully acquired at the youngest grade levels. However, a preliminary analysis showed that onset AoA was largely explained by the other item predictors (English: R^2 = .79; R^2 Spanish = .73). As a result, the variable was dropped from all models. From a substantive standpoint, this finding raises the possibility that gestural overlap might affect segmentability not only at the stage studied here (emerging phonemic awareness), but earlier on, during the acquisition of speech sounds for production, a topic worth further investigation.

5.5 General Discussion

This study set out to investigate the relationship between coarticulation and phoneme awareness. Unlike prior studies that have relied on phonological categories—namely [± sonorant] and [± continuant]—I used gestural overlap to proxy coarticulation. To date, no study has used this construct to predict segmentation performance. Data from this experiment's segmentation task resoundingly supported five out of six hypotheses; moreover, the only unexpected null finding (RQ6/H6) has a plausible articulatory explanation. I interpret this pattern of results as evidence that gestural overlap captures the essence of coarticulation in a way that [± sonorant] and [± continuant] have failed to do.

And in doing so, this investigation substantiates a long-held belief that, until now, has received only elusive empirical support: coarticulated speech undergirds phoneme awareness difficulty (Helfgott, 1978; Liberman et al., 1980; Treiman & Baron, 1981). Importantly, this relationship operates on a continuum, such that differences in degree of coarticulation translate to differences in segmentation difficulty. As Schreuder and van Bon (1989) hypothesized, it seems that the object of segmentation is indeed the articulatory code. Put another way: explicit phonemic awareness tasks[7] require operating on what one *says*, not simply what one *hears*—at least for beginning readers.

Such a conclusion has important implications in at least three areas: theory-building, instruction, and assessment.

First, theory development. If coarticulatory patterns cause segmentation difficulty, what mechanisms explain PA improvement? It has been widely reported that illiterate adults and

[7] For more on this distinction between explicit and implicit PA tasks, see Geudens and Sandra (2003).

children learning to read both struggle with phoneme-level tasks (Jiménez et al., 2007; Loureiro et al., 2004). Therefore, it seems reasonable to discard the possibility that developmental maturation is the primary driver of change. On the contrary, it seems that acquiring an alphabetic script plays an important role in shaping phonemic awareness (see Morais, 2021 for a review of the topic). Returning to the question, then: does PA improve because literacy acquisition refines phonological representations? If so, this would suggest that the object of segmentation changes according to literacy level. That is, a child who has reached a certain level of skill would rely on abstract representations (not the articulatory code) to execute PA tasks.

The above raises several interesting possibilities. First, might a side-effect of literacy be a change in articulatory programming and/or execution? Popescu and Noiray (2022) found that children who were good readers had measurably less coarticulated speech than those who were poor readers (see Saletta, 2015 for an overview of the relationship between literacy acquisition and changes in speech). Relatedly, should future models of lexical processing (cf., Dell, 1986) account for literacy knowledge? It seems reasonable to wonder whether refined phonological representations might change processing in some way. Consider lexical competition. If the target word is cat, words starting with the same sound and spelling (e.g., camel, caterpillar) may activate one another more than those starting with the same sound but a different spelling (e.g., kangaroo, chameleon).

Second, instruction. In the same way that a typical instructional sequence for PA places singletons before clusters, curricula should also consider between-cluster differences. Currently, materials tend to classify clusters by their second consonant (i.e., r- vs. l-clusters). Instead, results from this study imply three distinct classes of difficulty: fricative clusters, unvoiced-plosive clusters, and voiced-plosive clusters. This sequence of instruction might facilitate PA acquisition, especially for those who are most sensitive to the encoded articulatory information.

Another important consequence for instruction is that, precisely because coarticulation causes difficulty, beginning readers need a way to concretize these fleeting, amorphous spoken segments. Such ambiguity underscores the importance of learning the written representation (i.e., graphemes) of sounds. Put more directly: articulatory insights can and should be used to inform instructional sequencing and strategies, but the end goal of PA is reading and writing, both of which require knowledge of graphemes. Thus, teachers of beginning readers should be mindful that what a sound feels like or looks like does not always tell us how a word is spelled, and this phenomenon should be addressed with students.

Finally, assessment. First, measures of PA should include items of varied difficulty to ensure maximum coverage of the underlying participant ability distribution. Akin to the point made regarding instruction, results from the present investigation suggest that item categories should extend beyond simple and complex onsets. Test developers should specifically consider stimulus variation within complex onsets. An additional question from a psychometric standpoint is whether performance on singletons and clusters represent correlated but distinct subdimensions. Though PA has been generally described as a unidimensional construct (e.g., Schatschneider et al., 1999; Stanovich et al., 1984), the substantial range in difficulty between these item types leads me to wonder whether these differences are quantitative (*amount* of ability) or qualitative (*kind* of ability).

A more practical implication is the importance of using in-person, overt articulation during assessment. Historically, PA measures for educators (vs. researchers) were delivered in-person, one-on-one. Even if students were not asked to repeat stimuli before operating on them, the speaker's own gestures could have provided useful cues. Recently, however, there has been a rise in the use of pre-recorded stimuli delivered asynchronously via tablets or other devices. Especially if students are not prompted to repeat the stimuli, this change in modality has the potential to reduce access to important articulatory information.

5.6 Limitations

It is important to acknowledge several limitations of the preceding study. First and foremost, I was the sole investigator, which meant that all responses were transcribed and scored through my own biases and perceptions. Ideally, responses would have been processed by several individuals in order to estimate inter-rater reliability, especially for the most difficult-to-score items. A future study should examine data collected by various assessors who are blind to all research questions and hypotheses.

Another point worth highlighting is the study's testing protocol, whereby students repeated all stimuli before segmenting them. Geudens and Sandra (2003) recommended this procedure to ensure that students operate on the intended stimulus (p. 173). However, one could argue that inserting this step might have exacerbated the effect of gestural overlap. One response to this concern is that any augmentation would have been consistent across all stimuli. Thus, even if estimated difficulties are somehow exaggerated, the relative difference between onsets would remain.

A final and important limitation concerns the lack of discussion—and modeling of—the articulatory code's other dimensions beyond gestural overlap. Early studies have shown that spoken language is perceived both aurally and visually (McGurk & MacDonald, 1976; Sumby & Pollack, 1954); moreover, Castiglioni-Spalten and Ehri (2003) among others have shown that drawing children's attention to articulators can improve PA performance. Building on this idea, one intriguing avenue for future exploration is the degree to which "perceptual confusability" (Strand & Sommers, 2011, p. 16763) of phonemes and visemes (i.e., what our face looks like when producing a particular phoneme) influences phoneme awareness.

5.7 Conclusion

In sum, the present investigation uncovered persuasive evidence that coarticulation plays a decisive role in segmentation performance. It is the first study to show a direct link between the phonetic realization of speech and phoneme awareness. The inclusion of stimuli in two languages with a bilingual sample further solidifies the study's contribution to the field. Results indicate that segmentation behaviors have both cross-language commonalities and language-specific differences, nearly all of which could be predicted by previously documented coarticulatory patterns.

References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). *Fitting linear mixed-effects models using lme4.* Journal of Statistical Software, 67(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Blachman, B. A., Ball, E. W., Black, R. S., & Tangel, D. M. (1994). Kindergarten teachers develop phoneme awareness in low-income, inner-city classrooms. *Reading and Writing*, *6*(1), 1–18. https://doi.org/10.1007/BF01027275

Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206–244). Blackwell.

Bombien, L., & Hoole, P. (2013). Articulatory overlap as a function of voicing in French and German consonant clusters. *The Journal of the Acoustical Society of America*, *134*(1), 539–550. https://doi.org/10.1121/1.4807510

Bouwmeester, S., van Rijen, E. H. M., & Sijtsma, K. (2011). Understanding phoneme segmentation performance by analyzing abilities and word properties. *European Journal of Psychological Assessment*, *27*(2), 95–102. https://doi.org/10.1027/1015-5759/a000049

Bradley, T. G. (2007). Spanish complex onsets and the phonetics-phonology interface. *Optimality-theoretic studies in Spanish phonology*, 15-38. https://doi.org/10.1075/la.99.02bra

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology, 3*, 219–252. https://doi.org/10.1017/S0952675700000658

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*(2), 209–244. https://doi.org/10.1006/jpho.1996.0012

Castiglioni-Spalten, M. L., & Ehri, L. C. (2003). Phonemic awareness instruction: Contribution of articulatory segmentation to novice beginners' reading and spelling. *Scientific Studies of Reading, 7*(1), 25–52. https://doi.org/10.1207/S1532799XSSR0701_03

Catford, J. C. (1985). 'Rest' and 'Open transition' in a systemic phonology of English. In W. S. Greaves & J. D. Benson (Eds.), *Systemic perspectives on discourse* (Vol. 1, pp. 333–348). Normal, NJ: Ablex.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology: Volume 1: Between the grammar and physics of speech* (Vol. 1, pp. 283–333). Cambridge University Press. https://doi.org/10.1017/CBO9780511627736.017

Colantoni, L., & Steele, J. (2007). Voicing-dependent cluster simplification asymmetries in Spanish and French. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Current issues in linguistic theory* (Vol. 282, pp. 109–129). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.282.09col

Colantoni, L., & Steele, J. (2011). Synchronic evidence of a diachronic change: Voicing and duration in French and Spanish stop-liquid clusters. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *56*(2), 147–177. https://doi.org/10.1017/S0008413100003121

Content, A., Kolinsky, R., Morais, J., & Bertelson, P. (1986). Phonetic segmentation in prereaders: Effect of corrective information. *Journal of Experimental Child Psychology*, *42*(1), 49–72. https://doi.org/10.1016/0022-0965(86)90015-9

Crouch, C., Katsika, A., & Chitoran, I. (2023). Sonority sequencing and its relationship to articulatory timing in Georgian. *Journal of the International Phonetic Association*, *53*(3), 1049–1072. https://doi.org/10.1017/S0025100323000026

Cunningham, A. E. (1990). Explicit versus implicit instruction in phonemic awareness. *Journal of Experimental Child Psychology*, *50*(3), 429–444. https://doi.org/10.1016/0022-0965(90)90079-N

Cutting, J. (1975). Aspects of phonological fusion. *Journal of Experimental Psychology. Human Perception and Performance*, *104*, 105–120. https://doi.org/10.1037/0096-1523.1.2.105

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

De Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer. https://doi.org/10.1007/978-1-4757-3990-9

De Graaff, S., Hasselman, F., Bosman, A. M. T., & Verhoeven, L. (2008). Cognitive and linguistic constraints on phoneme isolation in Dutch kindergartners. *Learning and Instruction*, *18*(4), 391–403. https://doi.org/10.1016/j.learninstruc.2007.08.001

De Graaff, S., Hasselman, F., Verhoeven, L., & Bosman, A. M. T. (2011). Phonemic awareness in Dutch kindergartners: Effects of task, phoneme position, and phoneme class. *Learning and Instruction*, *21*(1), 163–173. https://doi.org/10.1016/j.learninstruc.2010.02.001

Duñabeitia, J. A., Cholin, J., Corral, J., Perea, M., & Carreiras, M. (2010). SYLLABARIUM: An online application for deriving complete statistics for Basque and Spanish orthographic syllables. *Behavior Research Methods*, *42*(1), 118–125. https://doi.org/10.3758/BRM.42.1.118

Ehri, L. C., Wilce, L. S., & Taylor, B. B. (1987). Children's categorization of short vowels in words and the influence of spellings. *Merrill-Palmer Quarterly*, *33*(3), 393–421.

Erbeli, F., Rice, M., Xu, Y., Bishop, M. E., & Goodrich, J. M. (2024). A meta-analysis on the optimal cumulative dosage of early phonemic awareness instruction. *Scientific Studies of Reading*, 1–26. https://doi.org/10.1080/10888438.2024.2309386

Geudens, A., Sandra, D., & Van den Broeck, W. (2004). Segmenting two-phoneme syllables: Developmental differences in relation with early reading skills. *Brain and Language*, *90*(1–3), 338–352. https://doi.org/10.1016/S0093-934X(03)00446-2

Geudens, A. & Sandra, Dominiek. (2003). Beyond implicit phonological knowledge: No support for an onset–rime structure in children's explicit phonological awareness. *Journal of Memory and Language*, *49*(2), 157–182. https://doi.org/10.1016/S0749-596X(03)00036-6

Gibson, M., María, A., Planas, A. M., Gafos, A., & Remirez, E. (2015). Consonant duration and VOT as a function of syllable complexity and voicing in a sub-set of Spanish clusters. *Interspeech 2015*.

Gibson, M., Bunta, F., Goodin-Mayeda, E., & Hernandez, A. (2018). The acquisition of syllable-level timing contrasts by English-and Spanish-speaking bilingual children with normal hearing and English-and Spanish-speaking bilingual children with cochlear implants. *Journal of Phonetics*, *78*, 98–112. https://doi.org/10.1016/j.wocn.2018.07.005

Gibson, M., Sotiropoulou, S., Tobin, S., & Gafos, A. I. (2019). Temporal aspects of word initial single consonants and consonants in clusters in Spanish. *Phonetica*, *76*(6), 448–478. https://doi.org/10.1159/000501508

Gonzalez, M. M. D. (2007). *Como detectar al niño con problemas del habla / How to detect a child with speech problems* (5th ed.). Editorial Trillas Sa De Cv.

Gu, Y. (2025). The CV coordination in English and Mandarin. *Proceedings of the Linguistic Society of America, 10*(1), 5916. https://doi.org/10.3765/plsa.v10i1.5916

Helfgott, J. A. (1976). Phonemic segmentation and blending skills of kindergarten children: Implications for beginning reading acquisition. *Contemporary Educational Psychology*, *1*(2), 157–169. https://doi.org/10.1016/0361-476X(76)90020-5

Jiménez, J. E. G., & García, C. R. H. (1995). Effects of word linguistic properties on phonological awareness in Spanish children. *Journal of Educational Psychology*, *87*, 193–201. https://doi.org/10.1037/0022-0663.87.2.193

Jiménez, J. E., Venegas, E., & García, E. (2007). Evaluación de la conciencia fonológica en niños y adultos iletrados: ¿es más relevante la tarea o la estructura silábica? *Infancia y Aprendizaje, 30*(1), 73–86. https://doi.org/10.1174/021037007779849691

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Wiley-Blackwell.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. https://doi.org/10.1037/h0020279

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*(2), 201–212. https://doi.org/10.1016/0022-0965(74)90101-5

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In H. Singer & R. B. Ruddell (Eds.), *Toward a psychology of reading* (Vol. 1, pp. 109–129). International Reading Association.

Liberman, I. Y., Liberman, A. M., Mattingly, I., & Shankweiler, D. (1980). Orthography and the Beginning Reader. In *Orthography, rending, and dyslexia.* University Park Press. https://haskinslabs.org/sites/default/files/files/Reprints/HL1084.pdf

Loevenbruck, H., Collins, M. J., Beckman, M. E., Krishnamurthy, A. K., & Ahalt, S. C. (1999). Temporal coordination of articulatory gestures in consonant clusters and sequences of consonants. In O. Fujimura, B. D. Joseph, & B. Palek (Eds.), *Proceedings of LP'98* (pp. 547–573). Karolinum Press.

Loureiro, C. de S., Willadino Braga, L., Souza, L. do N., Filho, G. N., Queiroz, E., & Dellatolas, G. (2004). Degree of illiteracy and phonological and metaphonological skills in unschooled adults. *Brain and Language, 89*(3), 499–502. https://doi.org/10.1016/j.bandl.2003.12.008

Marin, S., & Pouplier, M. (2010). Temporal Organization of Complex Onsets and Codas in American English: Testing the Predictions of a Gestural Coupling Model. *Motor Control*, *14*(3), 380–407. https://doi.org/10.1123/mcj.14.3.380

Marin, S., & Pouplier, M. (2014). Articulatory synergies in the temporal organization of liquid clusters in Romanian. *Journal of Phonetics, 42*, 24–36. https://doi.org/10.1016/j.wocn.2013.11.001

Massone, M. I. (1988). Estudio acústico y perceptivo de las consonantes nasales y líquidas del español. *Estudios de Fonética Experimental, 3*, 13–34.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

McBride-Chang, C. (1995). What is phonological awareness? *Journal of Educational Psychology*, *87*, 179–192. https://doi.org/10.1037/0022-0663.87.2.179

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*, 322–352. https://doi.org/10.1037/a0026744

Míguez-Álvarez, C. M., Cuevas-Alonso, M., & Saavedra, Á. (2021). Relationships between phonological awareness and reading in Spanish: A meta-analysis. *Language Learning*, *72*. https://doi.org/10.1111/lang.12471

Morais, J. (2021). The phoneme: A conceptual heritage from alphabetic literacy. *Cognition*, *213*, Article 104740. https://doi.org/10.1016/j.cognition.2021.104740

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, I. Ferrané, B. Vaxelaire, & R. Marsico (Eds.), *Approaches to Phonological Complexity* (pp. 299–328). De Gruyter Mouton.

Nam, H., & Saltzman, E. (2003). A competitive coupled oscillator model of syllable structure. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2253–2256.

Parker, S. G. (2002). *Quantifying the sonority hierarchy* (Doctoral dissertation, University of Massachusetts, Amherst). ProQuest Dissertations & Theses Global.

Parker, S. (2017). Sounding out sonority. *Language and Linguistics Compass*, *11*(9), e12248. https://doi.org/10.1111/lnc3.12248

Pater, J. (2009). Weighted constraints in Generative Linguistics. *Cognitive Science*, *33*(6), 999–1035. https://doi.org/10.1111/j.1551-6709.2009.01047.x

Popescu, A., & Noiray, A. (2022). Learning to read interacts with children's spoken language fluency. *Language Learning and Development*, *18*(2), 151–170. https://doi.org/10.1080/15475441.2021.1941032

Pouplier, M. (2020). Articulatory phonology. In M. Pouplier (Ed.), *Oxford research encyclopedia of linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.745

Pouplier, M., Pastätter, M., Hoole, P., Marin, S., Chitoran, I., Lentz, T. O., & Kochetov, A. (2022). Language- and cluster-specific effects in the timing of onset consonant sequences in seven languages. *Journal of Phonetics, 93*, Article 101153. https://doi.org/10.1016/j.wocn.2022.101153

Pouplier, M., Rodriquez, F., Lo, J. J. H., Alderton, R., Evans, B. G., Reinisch, E., & Carignan, C. (2024). Language-specific and individual variation in anticipatory nasal coarticulation: A comparative study of American English, French, and German. *Journal of Phonetics*, *107*, 101365. https://doi.org/10.1016/j.wocn.2024.101365

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rabe-Hesketh, S., & Skrondal, A. (2022). *Multilevel and longitudinal modeling using Stata* (4th ed.). Stata Press.

Ramírez, C. (2012). *Production and perception of the epenthetic vowel in obstruent + liquid clusters in Spanish: An analysis of the prosodic and phonetic cues used by L1 and L2 speakers* [Doctoral dissertation, University of Toronto]. https://tspace.library.utoronto.ca/handle/1807/32869

Rebernik, T., Jacobi, J., Jonkers, R., Noiray, A., & Wieling, M. (2021). A review of data collection practices using electromagnetic articulography. *Laboratory Phonology*, *12*(1). https://doi.org/10.5334/labphon.237

Rehfeld, D. M., Kirkpatrick, M., O, 'Guinn Nicole, & Renbarger, R. (2022). A meta-analysis of phonemic awareness instruction provided to children suspected of having a reading disability. *Language, Speech, and Hearing Services in Schools*, *53*(4), 1177–1201. https://doi.org/10.1044/2022_LSHSS-21-00160

Saletta, M. (2015). Literacy transforms speech production. *Frontiers in Psychology*, *6*, Article 1458.

Schatschneider, C., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91(*3), 439–451.

Schreuder, R., & van Bon, W. H. J. (1989). Phonemic analysis: Effects of word properties. *Journal of Research in Reading*, *12*(1), 59–78. https://doi.org/10.1111/j.1467-9817.1989.tb00303.x

Schröer, T. R. (2020). *Investigating potential acoustic correlates of sonority: Intensity vs. periodic energy* (Doctoral dissertation, Universität zu Köln).

Skjelfjord, V. J. (1987a). Phonemic segmentation. An important subskill in learning to read: I. *Scandinavian Journal of Educational Research*, *31*(1), 41–57. https://doi.org/10.1080/0031383870310104

Skjelfjord, V. J. (1987b). Phonemic segmentation. An important subskill in learning to read: II. *Scandinavian Journal of Educational Research*, *31*(2), 82–98. https://doi.org/10.1080/0031383870310203

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman and Hall/CRC. https://doi.org/10.1201/9780203489437

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *The Journal of Speech and Hearing Disorders*, *55*(4), 779–798. https://doi.org/10.1044/jshd.5504.779

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38*(2), 175–190. https://doi.org/10.1016/0022-0965(84)90120-6

StataCorp. (2023). *Stata 18 user's guide*. Stata Press. https://www.stata.com/bookstore/users-guide/

Strand, J. F., & Sommers, M. S. (2011). Sizing up the competition: Quantifying the influence of the mental lexicon on auditory and visual spoken word recognition. *The Journal of the Acoustical Society of America, 130*(3), 1663–1672. https://doi.org/10.1121/1.3613930

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215.

Solé, M.-J. (1992). Phonetic and Phonological Processes: The Case of Nasalization. *Language and Speech*, *35*(1–2), 29–43. https://doi.org/10.1177/002383099203500204

Solé, M.-J. (1995). Spatio-Temporal Patterns of Velopharyngeal Action in Phonetic and Phonological Nasalization. *Language and Speech*, *38*(1), 1–23. https://doi.org/10.1177/002383099503800101

Sotiropoulou, S., Gibson, M., & Gafos, A. (2020). Global organization in Spanish onsets. *Journal of Phonetics*, *82*, 100995. https://doi.org/10.1016/j.wocn.2020.100995

Navarro Tomás, T. (1932). *Manual de pronunciación española*. Consejo Superior de Investigaciones Científicas.

Treiman, R., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 3, pp. 159–198). Academic Press.

Treiman, R., & Weatherston, S. (1992). Effects of linguistic structure on children's ability to isolate initial consonants. *Journal of Educational Psychology*, *84*(2), 174–181.

Yavas, M. S., & Core, C. W. (2001). Phonemic awareness of coda consonants and sonority in bilingual children. *Clinical Linguistics & Phonetics*, *15*(1–2), 35–39. https://doi.org/10.3109/02699200109167627

Yavas, M. S., & Gogate, L. J. (1999). Phoneme awareness in children: A function of sonority. *Journal of Psycholinguistic Research*, *28*(3), 245–260. https://doi.org/10.1023/A:1023254114696

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, *5*(4), 683–689. https://doi.org/10.3758/BF03208845

Appendix A
Stimuli and Item Covariates

item_id	item_name	english	real_word	phonemes	vowel	onset_type	coda_cluster	log_freq
1	act	1	1	3	low	vc_obstruent	1	11.19
2	an	0	0	2	low	vc_sonorant	0	0.30
3	as	0	0	2	low	vc_obstruent	0	0.30
4	ate	1	1	2	diphthong	vc_obstruent	0	8.86
5	aus	0	0	2	diphthong	vc_obstruent	0	0.70
6	bleak	1	1	4	high	l_cluster_voiced_plosive	0	6.58
7	bli	0	0	3	high	l_cluster_voiced_plosive	0	0.30
8	blo	0	0	3	middle	l_cluster_voiced_plosive	0	0.48
9	blor	0	0	4	middle	l_cluster_voiced_plosive	0	0.60
10	bo	0	0	2	middle	singleton_voiced_plosive	0	0.70
11	bol	0	0	3	middle	singleton_voiced_plosive	0	1.11
12	bone	1	1	3	diphthong	singleton_voiced_plosive	0	9.68
13	bos	0	1	3	middle	singleton_voiced_plosive	0	1.15
14	bras	0	0	4	low	r_cluster_voiced_plosive	0	1.48
15	bro	0	0	3	middle	r_cluster_voiced_plosive	0	1.56
16	brought	1	1	4	middle	r_cluster_voiced_plosive	0	10.59
17	brusque	1	1	5	middle	r_cluster_voiced_plosive	1	4.16
18	clasp	1	1	5	low	l_cluster_unvoiced_plosive	1	6.01
19	class	1	1	4	low	l_cluster_unvoiced_plosive	0	11.45
20	claus	0	0	4	diphthong	l_cluster_unvoiced_plosive	0	1.65
21	clo	0	0	3	middle	l_cluster_unvoiced_plosive	0	1.81
22	co	0	0	2	middle	singleton_unvoiced_plosive	0	2.00
23	cons	0	0	4	middle	singleton_unvoiced_plosive	1	2.08
24	crept	1	1	5	middle	r_cluster_unvoiced_plosive	1	6.48
25	crest	1	1	5	middle	r_cluster_unvoiced_plosive	1	7.16
26	cro	0	0	3	middle	r_cluster_unvoiced_plosive	0	2.20

27	deck	1	1	3	middle	singleton_voiced_plosive	0	10.27
28	dice	1	1	3	diphthong	singleton_voiced_plosive	0	9.61
29	do	0	0	2	middle	singleton_voiced_plosive	0	2.23
30	don	0	1	3	middle	singleton_voiced_plosive	0	2.29
31	dro	0	0	3	middle	r_cluster_voiced_plosive	0	2.33
32	drum	1	1	4	middle	r_cluster_voiced_plosive	0	9.02
33	eel	1	1	2	high	vc_sonorant	0	7.79
34	eke	1	1	2	high	vc_obstruent	0	0.30
35	elf	1	1	3	middle	vc_sonorant	1	8.89
36	fan	1	1	3	low	singleton_fricative	0	10.53
37	flee	1	1	3	high	l_cluster_fricative	0	7.32
38	fli	0	0	3	high	l_cluster_fricative	0	2.36
39	flor	0	1	4	middle	l_cluster_fricative	0	2.41
40	flux	1	1	5	middle	l_cluster_fricative	1	8.63
41	fo	0	0	2	middle	singleton_fricative	0	2.46
42	foul	1	1	3	diphthong	singleton_fricative	0	8.31
43	fraught	1	1	4	middle	r_cluster_fricative	0	6.05
44	fri	0	0	3	high	r_cluster_fricative	0	2.50
45	fro	0	0	3	middle	r_cluster_fricative	0	2.51
46	glo	0	0	3	middle	l_cluster_voiced_plosive	0	2.58
47	go	0	0	2	middle	singleton_voiced_plosive	0	2.58
48	greek	1	1	4	high	r_cluster_voiced_plosive	0	9.90
49	gris	0	1	4	high	r_cluster_voiced_plosive	0	2.60
50	gro	0	0	3	middle	r_cluster_voiced_plosive	0	2.61
51	hitch	1	1	3	high	singleton_h	0	7.37
52	ic	0	0	2	high	vc_obstruent	0	2.62
53	ice	1	1	2	diphthong	vc_obstruent	0	10.55
54	if_it	1	1	2	high	vc_obstruent	0	14.65
55	il	0	0	2	high	vc_sonorant	0	2.67
56	ill	1	1	2	high	vc_sonorant	0	9.31

57	is	0	0	2	high	vc_obstruent	0	2.70
58	itch	1	1	2	high	vc_obstruent	0	6.63
59	jet	1	1	3	middle	singleton_fricative	0	9.42
60	kept	1	1	4	middle	singleton_unvoiced_plosive	1	10.56
61	key	1	1	2	high	singleton_unvoiced_plosive	0	11.47
62	ki	0	0	2	high	singleton_unvoiced_plosive	0	2.75
63	know	1	1	2	diphthong	singleton_sonorant	0	13.59
64	laus	0	0	3	diphthong	singleton_sonorant	0	2.77
65	leak	1	1	3	high	singleton_sonorant	0	8.55
66	lee	1	1	2	high	singleton_sonorant	0	10.57
67	li	0	0	2	high	singleton_sonorant	0	2.90
68	lo	0	1	2	middle	singleton_sonorant	0	2.93
69	mask	1	1	4	low	singleton_sonorant	1	9.42
70	na	0	0	2	low	singleton_sonorant	0	3.07
71	ño	0	0	2	middle	singleton_sonorant	0	3.08
72	nos	0	1	3	middle	singleton_sonorant	0	3.14
73	ol	0	0	2	middle	vc_sonorant	0	3.25
74	or	0	0	2	middle	vc_obstruent	0	3.26
75	ought	1	1	2	middle	vc_obstruent	0	9.63
76	own	1	1	2	diphthong	vc_sonorant	0	12.56
77	ox	1	1	3	middle	vc_obstruent	1	7.08
78	pent	1	1	4	middle	singleton_unvoiced_plosive	1	0.30
79	plea	1	1	3	high	l_cluster_unvoiced_plosive	0	7.59
80	pli	0	0	3	high	l_cluster_unvoiced_plosive	0	3.26
81	plo	0	0	3	middle	l_cluster_unvoiced_plosive	0	3.46
82	plon	0	0	4	middle	l_cluster_unvoiced_plosive	0	3.52
83	plump	1	1	5	middle	l_cluster_unvoiced_plosive	1	6.21
84	po	0	0	2	middle	singleton_unvoiced_plosive	0	3.58
85	prep	1	1	4	middle	r_cluster_unvoiced_plosive	0	7.69
86	price	1	1	4	diphthong	r_cluster_unvoiced_plosive	0	11.94

87	pro	0	0	3	middle	r_cluster_unvoiced_plosive	0	3.73
88	ras	0	0	3	low	singleton_sonorant	0	3.78
89	rate	1	1	3	diphthong	singleton_sonorant	0	11.38
90	raw	1	1	2	middle	singleton_sonorant	0	9.42
91	rep	1	1	3	middle	singleton_sonorant	0	9.32
92	rest	1	1	4	middle	singleton_sonorant	1	11.27
93	ro	0	0	2	middle	singleton_sonorant	0	3.85
94	rro	0	0	2	middle	singleton_sonorant	0	3.93
95	scuff	1	1	4	middle	s_cluster	0	5.78
96	self	1	1	4	middle	singleton_fricative	1	10.29
97	sell	1	1	3	middle	singleton_fricative	0	11.26
98	sigh	1	1	2	diphthong	singleton_fricative	0	9.05
99	so	0	0	2	middle	singleton_fricative	0	4.03
100	spent	1	1	5	middle	s_cluster	1	10.41
101	stone	1	1	4	diphthong	s_cluster	0	10.41
102	straight	1	1	5	diphthong	s_cluster	0	10.66
103	stud	1	1	4	middle	s_cluster	0	8.05
104	tact	1	1	4	low	singleton_unvoiced_plosive	1	6.63
105	threat	1	1	4	middle	r_cluster_fricative	0	9.83
106	to	0	0	2	middle	singleton_unvoiced_plosive	0	4.16
107	tone	1	1	3	diphthong	singleton_unvoiced_plosive	0	9.60
108	tor	0	0	3	middle	singleton_unvoiced_plosive	0	4.21
109	tract	1	1	5	low	r_cluster_unvoiced_plosive	1	7.42
110	trait	1	1	4	diphthong	r_cluster_unvoiced_plosive	0	7.64
111	tro	0	0	3	middle	r_cluster_unvoiced_plosive	0	4.24
112	tros	0	0	4	middle	r_cluster_unvoiced_plosive	0	4.25
113	weak	1	1	3	high	singleton_sonorant	0	9.75

Appendix B
Directions for Task Administration

English

"I'm going to say a word, and you're going to tell me all the sounds you hear. We'll do a few examples first."

Example stimuli: if, me, that, clip

Spanish

"Voy a decir una sílaba, y tú me vas a decir todos los sonidos en la sílaba. Algunas sílabas son palabras verdaderas, otras no. Vamos a practicar."

Example stimuli: nu, ec, tol, flar

Appendix C
Stata Code for Planned Comparisons

//Primary RQs (between language)

//RQ1:

```
lincom
((_b[5.onset_saturated#1.english]+_b[6.onset_saturated#1.english]+_b[7.onset_saturated#1.english]+_b[8.onset_saturated#1.english]+_b[9.onset_saturated#1.english]+_b[10.onset_saturated#1.english])/6)
```

//RQ2:

```
lincom _b[14.onset_saturated#1.english]
```

//RQ3:

```
lincom ((_b[1bn.onset_saturated#1.english] + _b[2.onset_saturated#1.english] +
_b[3.onset_saturated#1.english] + _b[4.onset_saturated#1.english])/4)
```

//Secondary RQs (within language)

//RQ4: Voicing of plosive clusters

```
lincom
((_b[6.onset_saturated]+_b[6.onset_saturated#1.english]+_b[8.onset_saturated]+_b[8.onset_saturated#1.english]))/2-
(_b[5.onset_saturated]+_b[5.onset_saturated#1.english]+_b[7.onset_saturated]+_b[7.onset_saturated#1.english])/2
```

```
lincom ((_b[6.onset_saturated]+_b[8.onset_saturated]))/2-
(_b[5.onset_saturated]+_b[7.onset_saturated])/2
```

//RQ5: Fricative clusters

```
lincom
((_b[9.onset_saturated]+_b[9.onset_saturated#1.english]+_b[10.onset_saturated]+_b[10.onset_saturated#1.english])/2-
(_b[5.onset_saturated]+_b[5.onset_saturated#1.english]+_b[7.onset_saturated]+_b[7.onset_saturated#1.english]+_b[6.onset_saturated]+_b[6.onset_saturated#1.english]+_b[8.onset_saturated]+_b[8.onset_saturated#1.english])/4)
```

lincom ((_b[9.onset_saturated]+_b[10.onset_saturated])/2-
(_b[5.onset_saturated]+_b[7.onset_saturated]+_b[6.onset_saturated]+_b[8.onset_saturated])/4)

//RQ6: l-clusters vs. r-clusters

* Including fricatives

lincom
((_b[7.onset_saturated]+_b[7.onset_saturated#1.english]+_b[8.onset_saturated]+_b[8.onset_saturated#1.english]+_b[10.onset_saturated]+_b[10.onset_saturated#1.english])/3-
(_b[5.onset_saturated]+_b[5.onset_saturated#1.english]+_b[6.onset_saturated]+_b[6.onset_saturated#1.english]+_b[9.onset_saturated]+_b[9.onset_saturated#1.english])/3)

lincom ((_b[7.onset_saturated]+_b[8.onset_saturated]+_b[10.onset_saturated])/3-
(_b[5.onset_saturated]+_b[6.onset_saturated]+_b[9.onset_saturated])/3)

* Excluding fricatives

lincom
((_b[7.onset_saturated]+_b[7.onset_saturated#1.english]+_b[8.onset_saturated]+_b[8.onset_saturated#1.english])/2-
(_b[5.onset_saturated]+_b[5.onset_saturated#1.english]+_b[6.onset_saturated]+_b[6.onset_saturated#1.english])/2)

lincom ((_b[7.onset_saturated]+_b[8.onset_saturated])/2-
(_b[5.onset_saturated]+_b[6.onset_saturated])/2)

Appendix D

Parameter Estimates from Model 4-Initial and Model 4-Unitized (logits)

	First Sound	**Unitized Sound**
	Est (*SE*) *p*	Est (*SE*) *p*
Onset Type		
Singleton unvoiced plosive	**2.45**	**-4.28**
	(0.34)	(0.35)
	< .001	< .001
Singleton voiced plosive	**2.14**	**-4.48**
	(0.36)	(0.38)
	< .001	< .001
Singleton fricative	**3.22**	**-4.78**
	(0.51)	(0.55)
	< .001	< .001
Singleton sonorant	**1.91**	**-4.17**
	(0.34)	(0.35)
	< .001	< .001
r-Cluster + unvoiced plosive	**0.85**	**-3.04**
	(0.37)	(0.38)
	0.02	< .001
r-Cluster + voiced plosive	0.17	**-2.58**
	(0.35)	(0.36)
	0.62	< .001
l-Cluster + unvoiced plosive	**0.74**	**-2.75**
	(0.36)	(0.37)
	0.04	< .001
l-Cluster + voiced plosive	0.07	**-2.43**
	(0.36)	(0.37)
	0.84	< .001
r-Cluster + fricative	**1.67**	**-3.91**
	(0.36)	(0.38)
	< .001	< .001
l-Cluster + fricative	**1.78**	**-3.73**
	(0.42)	(0.45)
	< .001	< .001
Vowel-initial + obstruent coda	**2.69**	**-4.71**
	(0.36)	(0.38)
	< .001	< .001
Vowel-initial + sonorant coda	**2.62**	**-5.21**

	(0.37)	(0.42)
	< .001	< .001
English × Onset Type		
Singleton unvoiced plosive	0.31	-0.61
	(0.36)	(0.44)
	0.38	0.16
Singleton voiced plosive	0.21	-0.75
	(0.35)	(0.47)
	0.55	0.11
Singleton fricative	-0.91	0.25
	(0.51)	(0.59)
	0.07	0.67
Singleton sonorant	0.42	-0.43
	(0.30)	(0.36)
	0.16	0.24
r-Cluster + unvoiced plosive	-0.27	**0.77**
	(0.32)	(0.37)
	0.41	0.04
r-Cluster + voiced plosive	**-0.74**	**1.77**
	(0.30)	(0.33)
	0.01	< .001
l-Cluster + unvoiced plosive	-0.25	0.60
	(0.33)	(0.37)
	0.44	0.11
l-Cluster + voiced plosive	-0.61	**0.94**
	(0.36)	(0.41)
	0.10	0.02
r-Cluster + fricative	**-1.13**	**1.03**
	(0.35)	(0.41)
	< .001	0.01
l-Cluster + fricative	-0.67	0.53
	(0.39)	(0.44)
	0.08	0.22
Vowel-initial + obstruent coda	**-0.79**	-0.40
	(0.34)	(0.42)
	0.02	0.35
Vowel-initial + sonorant coda	**-1.80**	**1.11**
	(0.37)	(0.48)
	< .001	0.02
Vowel Type		
Middle	**-0.22**	**0.34**
	(0.10)	(0.12)

	0.03	< .001
Low	**-0.35**	**0.54**
	(0.14)	(0.17)
	0.01	< .001
Diphthong	**-0.36**	**0.57**
	(0.13)	(0.17)
	0.01	< .001
Coda Cluster	**-0.40**	0.14
	(0.18)	(0.22)
	0.03	0.52
Phonemes	-0.00	-0.14
	(0.11)	(0.13)
	0.98	0.26
Log_Freq	-0.07	**-0.11**
	(0.04)	(0.05)
	0.09	0.03
Real Word	0.31	0.11
	(0.24)	(0.26)
	0.19	0.68
Spanish at home	-0.26	0.09
	(0.19)	(0.20)
	0.19	0.65
Program Type		
90% English	0.21	**0.95**
	(0.30)	(0.27)
	0.48	< .001
>= 50% Spanish	0.20	0.52
	(0.34)	(0.31)
	0.56	0.09
Grade		
Grade 1	**1.28**	-0.46
	(0.31)	(0.27)
	< .001	0.09
Grade 2	**1.71**	**-0.55**
	(0.31)	(0.27)
	< .001	0.04

TK	**-1.89**	1.09
	(0.88)	(0.67)
	0.03	0.11
Random Part		
$\psi_4^{(3)}$	0.49	0.20
	(0.18)	(0.13)
$\psi_{22}^{(2)}$	2.52	2.71
	(0.34)	(0.46)
$\psi_{11}^{(2)}$	2.04	1.99
	(0.27)	(0.31)
$\psi_{21}^{(2)}$	**1.67**	**1.68**
	(0.24)	(0.29)
	< .001	< .001
Log-likelihood	-4152.7	-3089.3
Number of observations	12,388	12,388