

Variation in syllabification: onglides in Sonoran Spanish

Abstract

Postconsonantal glides (*m[je]do* ‘fear’) have traditionally been considered part of the nucleus in Spanish. Recent work, however, has challenged nuclear affiliation. This research examines the syllabic affiliation of postconsonantal glides in Sonoran Spanish to determine if they are always parsed in the nucleus or they can at least at times be parsed in the onset. Experimental data were collected using a phrase reading and a syllable division task. Stimuli were designed according to two hypotheses. The first relied on the generalization that Spanish allows a maximum of three rhyme segments. If subjects produce CGVGC sequences, the postconsonantal glide should be in the onset, as otherwise the rhyme would contain four segments and be illicit. The second hypothesis predicts that, if the glide is in the onset, there should be onset co-occurrence restrictions.

The study findings showed that glides can be *variably* parsed in a complex onset. Sonoran speakers produced the GVG sequence at least some of the time and thus it is possible for it to be parsed as the second segment in a complex onset. Acoustic analysis revealed that pre-vocalic segments are longer after palatal consonants and the duration increase was due to lengthening of the onset (a strategy to avoid an illicit sequence). Onset glides in Sonoran Spanish provide evidence for cross-dialectal and intra-dialectal variation in syllabic affiliation. It is shown that an optimality theoretic model can capture and explain the sources of internal variation as competing factors which result in variation in output parsing. It is concluded that the affiliation debate has been oversimplified by considering either an onset or a nuclear parse.

Keywords: glides, onset, nucleus, syllabification,

1. Introduction

Overall, Spanish syllabification is uncontroversial. Native speakers tend to have clear intuitions about how to divide words into syllables. And yet, despite general agreement, areas of debate have emerged in recent decades suggesting that Spanish syllabification may not be as straightforward as it is generally believed. Examples of controversial topics are the syllabification of vowel sequences as diphthongs or hiatuses and the syllabic affiliation of prevocalic glides (i.e., onglides). Some sequences of high vowel and another vowel, which are normally syllabified in one syllable as a diphthong, surface as hiatuses in some Spanish dialects (Hualde 1997, 1999, 2002; Colina 2009). The syllabic affiliation and moraic status of vocoids (in diphthongs or hiatuses) is relevant because it has implications for the phonemic or allophonic status of glides, the nature of linguistic competence, and the mental representation of exceptional patterns. Another issue under debate involves postconsonantal glides, which have traditionally been considered part of the nucleus. Recent work, however, has challenged nuclear affiliation and argued for an onset parsing or for the possibility of both parsings.

This paper will focus on the syllabic affiliation of onglides, in particular of those in postconsonantal position, that is glides preceded by a consonant or more in the same syllable.

After a review of the literature and the arguments for onset and nuclear parsing, we argue that the debate has been oversimplified by considering either an onset or a nuclear

parse. We will challenge that position by showing that, although there may be preference for a nuclear affiliation, an onset parse is also possible in some dialects. This is supported by the results of an experimental study that indicate that in some Southwest Spanish varieties postconsonantal glides can go in the onset to avoid co-occurrence restrictions and limits in the number of nuclear segments. Additionally, we argue that an optimality theoretic model can capture and help understand the sources of language internal and external variation as competing factors which result in variation in output parsing.

2. Syllabic affiliation of onglides: onset or nucleus?

It is not surprising that glides are involved in controversial aspects of the syllabification of Spanish. Glides are opportunistic segments because of their featural composition [-consonantal, -vocalic], which makes them neither vocalic nor consonantal, and of their ability to be parsed in the margins of the syllable. It is thus their hybrid nature that makes their syllabic affiliation unclear.

In this section we provide a critical review of the affiliation arguments for postconsonantal onglides; first for a nuclear parsing, as it reflects the most common position among phonologists, followed by arguments for an onset affiliation.

The arguments in favor of the nuclear status of prevocalic postconsonantal glides have to do with co-occurrence restrictions, rhyme restrictions, diphthong/monophthong alternations, hypochoristic formation, vowel harmony, children's games, acquisition, intervocalic glides, and stress (2.1-2.8 below) (cf. Hualde 2014, Shelton et al. 2012, Colina 2009).

2.1. Co-occurrence restrictions. Co-occurrence restrictions are often used as evidence that two segments cannot be part of the same subsyllabic constituent, whereas lack thereof is an indication that two contiguous segments are parsed in two separate constituents. For instance, complex onsets in Spanish have to exhibit a minimal sonority distance often described as that between an obstruent, the least sonorous of the consonants, and a liquid, the most sonorous. Therefore, a combination of a sonorant plus sonorant in an onset cluster, such as *[ml], *[nl], *[mr], *[nr], *[nw] would be ill-formed. That sonorant- plus-glide sequences such as [mj], [nw] [rwV] [lwV] (*miedo* ‘fear’, *nueve* ‘nine’, *rueda* ‘wheel’, *liebre* ‘hare’) are possible sequences in Spanish has been used as evidence that [j] and [w] must be in the nucleus (Shelton et al. 2012). This is in contrast with ill-formed *[nw], *[rwV], and *[lwV] in English that are thus argued to be in the onset.

Co-occurrence restrictions can also involve features such as point of articulation and voicing. In Spanish *[tl] is ill-formed as an onset cluster in most dialects and *[dl] in all, and this is attributed to a prohibition against homorganic segments in the onset (coronal for *[tl]) and also identical voicing specification in the case of *[dl] [Harris 1983]). An argument for a nuclear parsing of glides based on point-of-articulation co-occurrence restrictions rests on the well-formed status of a sequence of a bilabial stop and labial glide, such as [bw] in [bwé.no] ‘good’. Languages that parse this sequence in the onset such as English do not allow [*bw] because both segments are labial and in the onset. Consequently, the argument is that since in Spanish [bweno] is well-formed, the glide must be parsed in the nucleus (Shelton et al. 2012).

Martínez-Gil (2016: 156), however, argues for the opposite position—glides are parsed as part of a complex onset—on the grounds that there exist co-occurrence restrictions in

point of articulation between a palatal and a high front glide. Thus, while a palatal lateral or fricative can be followed [w] as in [jweβe] ‘it rains’ or [poɫwelo] ‘chick’, they are ill-formed if followed by a [j] front glide because it shares the same point of articulation [+high], as in the hypothetical *[jjeβe] *[poɫjelo].

While it is possible to question the extent of the restrictions on palatals on the basis that they could be lexically restricted or due to historical reasons (palatal laterals in Spanish derive from complex segments, e.g. [lj], geminates, etc.), it is also reasonable to challenge Shelton et al.’s argument that glides are nuclear because a sonorant + glide onset cluster would violate sonority restrictions. While that is true, sequences in which the consonant preceding the glide is not a sonorant, such as an obstruent, abide by minimal sonority distance and thus the glide could potentially be in the onset.

The number of permissible onset segments serves as the basis of another argument for nuclear parsing of glides in Spanish. Many languages, including Spanish and English, limit the number of onset segments to two (unless one is /s/ in the case of English, which then permits three onset segments). Consequently, the postconsonantal glides in [trweno] ‘thunder’ and [pljejo] ‘fold’ in Spanish must be nuclear because an onset parse would result in a three-member cluster. However, this argument does not rule out the possibility of an onset glide when the onset has only one constituent and it is not homorganic (as in the experiment described in Section 4).

A final phonotactic argument for a nuclear parsing consists of contrasts such as *escuela* ‘school’, with an epenthetic [e] before [s], vs. *sierra* ‘mountain range’, where epenthesis is not needed (**esierra*). Shelton et al. (2012) interpret this to mean that the high glide after [s] in *sierra* must be in the nucleus, and therefore it does not have an effect on the

composition of the onset. Although one cannot argue with nuclear affiliation of the glide in *sierra*, the strength of the argument for a nuclear parse of glides can be questioned because of contextual differences in the two examples. An epenthetic [e] is inserted in *escuela* to repair a decrease in sonority (from [s] to [t]) in a potential onset cluster; however, this issue is not at stake in *sierra* where sonority rises from [s] to [j]. As seen in the case of co-occurrence restrictions on onset clusters, where sonorant + glide cannot be in the onset, but an obstruent + glide could, it sometimes is the case that a restriction and its derived argument apply only to a subset of segments or to a specific context, and, therefore, cannot be used as evidence against onset glides in general.

2.2. Diphthong/monophthong alternations. Spanish has glide-plus-vowel sequences that alternate with monophthongs in unstressed morphologically related forms such as *cuello* 'neck', *collar* 'necklace', *ciego* 'blind person', *ceguera* 'blindness'. This fact is presented as evidence for a nuclear affiliation of the glide since it is a realization of a monophthong that can be said to have a tighter connection with the nucleus than with the onset (Shelton et al. 2012). Martínez-Gil (2016) counters this argument by pointing out that some of these alternations result in an onset glide, as demonstrated by their consonantization, as in *helar* and *hielo* [jelo]. However, it should be noted that, as in some of the arguments summarized above, the phonological contexts are not the same in these two sets of examples: the glide/fricative in *hielo* [jelo] is not postconsonantal as it is in *cuello* and *ciego* and therefore the glide could be in the nucleus in *ciego* and in the onset in [jelo].

2.3. Hypochoistics. Hypochoistics such as *Dani* and *Javi*, from *Dan[je]l* and *Jav[je]r* respectively, are presented as evidence for the nuclear affiliation of the glide because it appears as a full vowel in the hypochoistic (Prieto 1992; Colina 2009; Shelton et al. 2012, Hualde 2014), alternating thus with a full vowel. If it were in the onset, an additional vowel would be required, as can be seen in examples such as **Petr*, *Petro* from *Petronio*. Nonetheless, Martínez-Gil (2016) contends that this cannot be considered definitive evidence for a nuclear parsing since forms like *Loli* (from *Dolores*, *Lola*) and *Pili* (from *Pilar*) indicate that [i] can form hypochoistics independently of the presence of a diphthong in the base.

2.4. Vowel harmony: Some dialects of Northwestern Spain have a high harmony process in which a stressed high vowel raises all unstressed vowels to its left (Hualde 1991). In these varieties a prevocalic high glide behaves like a high vowel in that it triggers high harmony, as *mi lu djó* ‘he/she gave it to me’ vs. *me lo compró* ‘he/she bought it for me’ and can therefore be argued to be in the nucleus. This type of evidence is considered weak by some as it refers to a minority dialect seen as distantly related to general varieties.

2.5. Children’s games: Data from children’s games have been used to support the nuclear affiliation of prevocalic glides. Hualde (2014: 199) refers to a game of *jerigonza* in which [k] is inserted after each vowel in a word and followed by a copy of the vowel, for instance, *pasa* ‘it happens’ becomes *pa-ka-sa-ka* and *están* ‘they are’ is *e-kes-ta-kan*. In this game a second member of an onset cluster is not copied, however a prevocalic glide is, as seen in *proclama* ‘it proclaims’ *pro-ko-cla-ka-ma-ka* vs. *puente* ‘bridge’ *pu-ku-e-ken-te-ke* or *tiene*

‘it has’ *ti-ki-e-ke-ne-ke*. Yet, this is unlike other varieties, which insert [p] and omit the prevocalic glide, *canCIÓN* ‘song’ *cam.pa.cióm.po* (Piñeros 1998: 61) indicating that the glide behaves like an onset consonant. Shelton et al. (2012: 331) mention additional types of *jerigonza* in which off glides and codas are maintained, but onsets and onset glides are not: *estoy* ‘I am’ *e-pes-to-poy*, but *también* ‘also’ *ta-pam-bie-pen* and *puerta* ‘door’ *pue-per-ta-pa* also supporting the position that prevocalic glides behave like onsets and are not nuclear. While the evidence seems contradictory, it is important to notice that it is drawn from different dialectal varieties which could exhibit differences in syllabification. This is a noticeable aspect in the literature that reviews the arguments for nuclear or onset parsings, which often does not discriminate between dialects, combining data from various dialects.

2.6. Acquisitional arguments: Studies of children with phonological delay, who simplify onset clusters to singletons, offer information regarding the syllabic affiliation of onglides. Researchers hypothesized that if CGV sequences are structurally related to CLV sequences then treatment with one group would generalize to the other. Anderson (2002) found that treatment of CGV sequences resulted in improvement of CLV sequences, suggesting that glides and liquids are syllabified in the same manner, as onsets. Other studies, however, report the opposite findings. Barlow (2005), for instance, indicates that performance on CLV sequences improved following treatment but performance on CGV sequences did not. In support of an onset parse, Kehoe et al. (2008) report that the production of rising diphthongs and branching onsets patterned similarly for the children in their study. Overall, these findings suggest variation in parsing and that, at least for some children, during the early stages of acquisition prevocalic glides can be part of a branching onset.

2.7. Intervocalic glides. Glides in intervocalic position (when not preceded by a consonantal onset) become consonantal in most varieties of Spanish, with realizations that range from a fricative to a stop or affricate, e.g. [-jendo] *comiendo*, [ko.mjen.do] vs. *creyendo* [kre.jen.do]. This fact has been referred to by some (Martínez-Gil 2016) as evidence that glides are in the onset; others (Colina 2009, Hualde 2005) have taken the opposite position, contending that glide consonantization is proof that prevocalic glides are not possible onsets (unless they become consonantal) and must therefore be nuclear. The contrast in argumentation is reflective of a difference in theoretical assumptions. Martínez-Gil (2016) is framing the matter in a serial understanding of phonology in which the glide must be in the onset to become a consonant, while Hualde (2005) and Colina (2009) conceptualize the issue in a parallel framework in which consonantization is a repair mechanism to avoid an ill-formed glide in onset position. In other words, Martínez-Gil's argument presupposes a derivational account in which a glide is syllabified in the onset at some point in the derivation, and then becomes an obstruent. The glide, however, never surfaces in the onset in these dialects, weakening support for onset glides. We will return to this argument in Section 3.

Once again context is shown to have a confounding effect on the debate over the syllabic affiliation of onset glides, as the pertinent position is not intervocalic, rather postconsonantal and prevocalic: this is the position where a glide can be potentially parsed in a complex onset or a complex nucleus, as in *comiendo*, [ko.mjen.do].

2.8. Stress. Stress facts have been brought to bear on the glide debate. Spanish stress can fall on the last, penultimate or antepenultimate syllable (Harris 1983; Morales-Front 2014). This three-syllable window for stress placement is however reduced to two when the penultimate contains a prevocalic glide or a coda, a fact that has been presented as evidence that the prevocalic glide must be part of the rhyme (i.e., nucleus) because it counts for stress purposes, **Venézwela Venezwéla*. Nonetheless, in what runs counter to a nuclear parse for the prevocalic glide, the stress data also support a difference in behavior between prevocalic and postvocalic glides: a postvocalic glide in final position always attracts stress, e.g., *convóy*, ‘convoy’, while this is not the case for a prevocalic glide in the same position, *família*, ‘family’ vs. *Meliá*. The contradictory evidence derived from stress facts becomes irrelevant for the glide debate if, as Hualde (2005) does, the above generalizations are interpreted not as the consequence of a difference in parsing of the glides, but rather as a historical remnant from Latin. Spanish inherited stress in the same position it had in Latin. The unmarked antepenultimate stress of Latin became penultimate if the penultimate had a long vowel or a closed syllable, which explains the preference for the reduction of the syllable window when the penultimate has a heavy syllable. Latin also had no rising diphthongs like Spanish, which were pronounced in hiatus. This explains why forms like *família* are possible. In sum, under the historical explanation of the stress patterns in Spanish, the stress data do not offer solid evidence for or against the nuclear parsing of glides, as the different behavior of glides with regard to stress can be attributed to a historical fact.

2.9. Experimental evidence on stress and diphthongs also supports a difference in behavior between on and off glides, which in turn are different from monophthongs. Shelton (2007), Shelton et al. (2010) and Shelton et al. (2012) found that speakers who were asked to pronounce nonce words made more errors in antepenultimately stressed syllables when the penultimate contained a falling diphthong than with a rising one, and rising diphthongs had in turn more mistakes than monophthongs (e.g. *loteiga* > *lotiega* > *lotega*). These differences can be ascribed to the position of the glide with respect to other components of the syllable, as a postconsonantal glide can move into an unoccupied onset or next to a singleton onset as the second member of a cluster, whereas a postvocalic glide can only move into a nuclear position, remaining in the rhyme (cf. as proposed also by Colina and Simonet 2014, for Galician coda clusters, who argue that a nasal glide is parsed in the nucleus to avoid a complex coda clusters).

In summary, despite a predominance of phonological argumentation favoring a nuclear affiliation of postconsonantal glides, some evidence suggests that an onset parsing is also possible. We argue that the debate in the literature has been oversimplified by examining the arguments for one position or the other and trying to resolve the matter in favor of either an onset or a nuclear parse, rather than allowing for both possibilities under different conditions (cf. Shelton et al. 2010, Hualde 2014, Martínez Gil 2016, etc.). As mentioned above, an onset parse can be an option for preconsonantal glides to avoid ill-formed configurations incurred by a nuclear parse. For instance, if the nucleus consists of a vowel preceded by a glide and followed by a glide and a coda consonant, an onset parse of the first glide may help avoid having three segments in the nucleus. Similarly, dialectal variation is also possible for syllabification, and some varieties of Spanish may prefer a

complex onset consisting of consonant and glide to a complex nucleus, as will be seen below. Finally, some variation may be due to lexical and postlexical distinctions. This is what Martínez-Gil (2000, 2016) proposes for syllable merger in Chicano Spanish, in whose analysis prevocalic glides and demorification of the first vocoid in a sequence is said to be the result of an onset parse postlexically.

Variation in syllabic affiliation, although not frequently considered, should not be entirely unexpected, especially in view of the non-phonemic, predictable status of syllabification in Spanish. We argue that prevocalic and postvocalic glides are opportunistic in ways not yet fully considered, in that their parsing can fluctuate depending on phonological context (such as the composition of the onset preceding it and the rhyme following, *ciego*, *jelo*), dialectal variation, and lexical and postlexical considerations, thus explaining the lack of clear-cut categorical intuition on syllabic affiliation. While Shelton et al. (2012) propose that this “inconsistent patterning of on-glides reported in the various kinds of literature... might also be best interpreted from a probabilistic perspective (341)”, other frameworks such as optimality-theory are ideally suited to account for conflicting constraints that produce different syllabic affiliation of glides.

3. Glides in Sonoran Spanish

In this paper we investigate an example of variation in the syllabic affiliation of postconsonantal, prevocalic glides, to determine whether the glide can be in the onset.

As mentioned above, in most varieties of Spanish, when the glide is prevocalic, but not postconsonantal (i.e., the onset is empty), the glide becomes a consonant, generally an

obstruent (with various degrees of aperture/constriction, e.g., approximant, fricative, stop, affricate) and it is parsed in the onset, e.g., *-iendo* [-jen.do] ‘-ing’, *com-iendo* [ko.mjen.do] ‘eating’, vs. *creyendo* [kre.jen.do] ‘believing’ (Hualde 2005, Colina 2009). In other words, glides are not possible onsets. The ill-formedness of glides in the onset is one of the arguments used to support the nuclear parsing of prevocalic glides (Hualde 2005, Colina 2009). However, some Spanish varieties, such as the Spanish of Sonora, Mexico, allow glides in the onset (Canfield 1981, Alvar 1996), i.e., *creyendo* ‘believing’ [kre.jen.do], which do not consonantize. Consequently, if glides are well-formed single onsets, a glide preceded by a consonant (e.g., *i* in *tiara*) could in principle be parsed as the second element in the onset [tj] (i.e., an onset cluster), rather than in the nucleus [ja] given that it conforms to sonority restriction on onset clusters. In Section 4 we describe two experiments that set out to test this hypothesis regarding the parsing of glides in an onset cluster.

Section 2 reviewed the literature on syllabic affiliation of glides. The following paragraphs focus specifically on proposals for glide affiliation in Sonoran and more broadly on Southwest Spanish. Southwest Spanish is a variety of Spanish spoken in the southwest of the United States, in states such as Texas, Arizona, and California. Some authors also refer to it as Chicano Spanish (Martínez-Gil 2000, 2016). Southwest Spanish encompasses the variety of Sonoran Spanish spoken in border regions of Southern Arizona and in the state of Sonora, Mexico. Glides in these varieties have received significant attention in the context of across-the-word vowel merger (Hutchison 1974, Clements and Kayser 1980, Martínez-Gil 2000, 2016, Bakovic 2006, Colina 2009) and are relevant to the debate on glide affiliation, to the phonotactics of high vocoids and to dialectal variation in phonotactics.

In Southwest Spanish vowel merger, postconsonantal, prevocalic high and mid vowels surface as high glides (1a), and are deleted when they agree in backness with the following vowel (1b). A low vowel is deleted in the same context (1c). The vowel affected is always the first one in a sequence.

(1) Vowel merger in Southwest Spanish vs. Peninsular Spanish

	Southwest Spanish	Peninsular Spanish	Gloss
a. me iría	[mi.rí.a]	[mej.ría]	'I would go'
b. me usó	[mju.só]	[m̠eu.só]	'S/he used me'
c. habla inglés	[a.βliŋ.glés]	[a.βlajŋ.glés]	'S/he speaks English'

Colina (2009) interprets this as a sonority-based preference for high glides (over mid) in this Spanish variety (in contrast with others like Peninsular Spanish), and for a complex nucleus rather than a complex onset. Like Bakovic (2006), Colina (2009) assumes the then standard position (and standard arguments) for the nuclear affiliation of glides (Colina 2009: 21; Colina 2012: 141-142). Martínez-Gil (2000, 2016), however, argues for an onset position, in which it is the onset parsing of the first vowel that drives the ban on mid vowels as well as the selection of the first vowel as the target for gliding and demorification. In other words, Martínez-Gil's explanation of the dialectal differences in vowel merger is that Sonoran Spanish repairs an onsetless syllable (the one in word-initial position) through the onset parsing of the first vocoid, while other accounts claim that this is done through the creation of a complex nucleus. A robust argument in favor of Martínez-Gil's proposal is that onset affiliation explains why it is the first vowel, not the second, that glides. Colina

(2009: 59-65) points to a preference for a complex nucleus over a coda (domination of the constraint *Coda over *Complex Nucleus); yet no mention is made in the analysis of an alternative parse which violates *Complex Onset, because this possibility was ruled out through the arguments in favor of the nuclear parse.

Another piece of evidence in favor of onset parsing of glides in Sonoran Spanish presented by Martínez-Gil relates to the lexical/postlexical distinction. He states that, although arguments for lexical syllabification of postconsonantal glides favor the nucleus in many dialects, this does not affect the postlexical level. Spanish varieties that ban high glides in onset position allow them, however, when they are the result of across the word resyllabification (2). In derivational terms, onset glides are ill-formed lexically in these dialects, but they are permitted postlexically. As Martínez-Gil (2000, 2016) indicates, vowel merger in Southwest Spanish is a postlexical process which applies across words.

(2) ley [lej] leyes [le.jes] ley alguna [le.jal.yu. na]

A final argument in support of Martínez-Gil's analysis is that Sonoran Spanish has glides in singleton onset positions, that is, they do not consonantize as in other dialects. As discussed in Section 2, Martínez-Gil (2016) resorts to consonantization to support his view of an onset parsing for the glide, indicating that, because the glide is strengthened, it must be in the onset. The reader will recall the objection that this argument is framework dependent and that it only shows that glides are not well-formed in the onset (at least lexically), and therefore must become consonants. Nonetheless, Martínez-Gil's (2016) consonantization argument for onset affiliation of glides turns out to be unnecessary as Sonoran Spanish does in fact allow onset glides without consonantization when they are the only segment in the onset. Whether glides are allowed as the second member of an

onset cluster remains to be investigated and it is the goal of the experiment described in this paper.

4. Experiment

4.1 Research questions

We set out to test the hypothesis that, since onset glides are well-formed single onsets in the Spanish of Sonora, Mexico, a glide preceded by a consonant (e.g., *i* in *tiara*) could be parsed as the second element in the onset [tj] (i.e., an onset cluster), rather than in the nucleus [ja] because it conforms to sonority restrictions on onset clusters. The research question (RQ) is thus the following:

RQ: Can a postconsonantal, prevocalic glide be parsed as the second segment of a complex onset in Sonoran Spanish? In other words, are Sonoran Spanish prevocalic glides always part of a complex nucleus or can the glide be parsed in the onset, at least sometimes?

A positive answer to the RQ would support the position (presented in Section 2) that the debate in the literature has been oversimplified by trying to decide in favor of either an onset or a nuclear parse for Spanish postvocalic glides, rather than considering the possibility of onset as well as nuclear parses, under different conditions.

In addition, onset parsing would suggest cross-dialectal variation in the syllabic affiliation of prevocalic glides. It is reasonable to expect that dialectal variation will affect syllabic affiliation, but little is known about cross-dialectal variation in this area of Spanish syllabification. Onset parsing of the prevocalic glide would bear out the predictions of an optimality-theoretic factorial typology in which the effects of a higher ranked constraint

against a super heavy nucleus can force a nucleus glide into the onset rather than the nucleus (as an alternative, vs. an either or situation). We will return to this in the Discussion.

4.2 Procedure, participants and materials

Oral data were recorded in a sound isolated booth using a Shure SM10A Head mounted microphone. The participants were 10 speakers of Sonoran Mexican Spanish, 18-25 years old. They were asked to perform two tasks: (i) an oral syllable division task and (ii) an oral phrase reading in which the stimulus was provided in a sentence (e.g., *Digo ____ porque sí* 'I say ____ because I say so'). The stimuli consisted of a total of 21 nonce words (alongside 21 fillers) with a C+G+V sequence.

Two sets of stimuli were designed according to two hypotheses.

Set 1, Hypothesis 1:

Spanish only allows a maximum of three rhyme segments (Harris 1983). If a sequence of CGVGC (Consonant + Glide + Vowel + Glide + Consonant) is allowed, the glide should be in the onset, because otherwise the rhyme would contain four segments and would be illicit. According to this, a stimulus set was created that consisted of 14 four-syllable nonce words containing a postconsonantal high vocoid followed by a diphthong and a coda consonant. CGVGC monosyllabic parsings (which do not divide the sequence in more than one syllable) will indicate that the glide can be parsed in the onset, as they must consist of a complex onset + three segment rhyme, e.g., *lacap[jaj]sto*, *lacap[waj]sto* because a four-segment rhyme would be ill-formed (See Appendix). Four syllable words were created to avoid a glide + vowel sequence too close to the beginning or end of the word, a position known to favor hiatuses in some dialects (Hualde 1999, 2005).

Set 2, Hypothesis 2:

If the glide is in the onset, there will be onset co-occurrence restrictions, i.e., only some combinations of consonant + glide should be possible as complex clusters. Crucially, a palatal consonant + homorganic glide [j] (e.g., **ch[j]aba*, *[j]ape, **ma[nj]ala*) should be disallowed because their articulations are too similar; in that case, one can argue that the glide is in the onset. On the basis of this hypothesis, 7 nonce words (not controlled for number of syllables or stress, all with the same G and V for GV, *ia*) were used as stimuli.

4.3. Statistical analyses

We report three primary statistical analyses in order to address the aforementioned research question. To this end, we fit a series of Bayesian regression models, which are described in detail in the corresponding sections below and in the supplementary materials (https://osf.io/fyt4d/?view_only=1e5f867b1896417cbaa21a2872221cf8). All analyses were conducted in R (R Core Team, 2018, version 4.1.0). The models were fit using *stan* (Stan Development Team, 2018) via the R package *brms* (Bürkner, 2017). All models included a maximal grouping-effects structure (Barr, Levy, Scheepers, & Tily, 2013), as well as regularizing, weakly informative priors (Gelman, Simpson, & Betancourt, 2017). Additionally, models were fit with 2000 iterations (1000 warm-up) and Hamiltonian Monte-Carlo sampling was carried out with 4 chains distributed between 4 processing cores. We report point estimates (posterior medians) for each parameter of interest, along with the 95% highest density interval (HDI), and the maximum probability of effect (MPE).

A complete description of the models and output summaries in table format are available in the supplementary materials.

4.4 Results

4.4.1 Syllable division task

The first model analyzed the syllable division task data using hierarchical multinomial logistic regression. The participants responses to the critical CGVG sequences were classified as triphthong, hiatus, or simplification, i.e., [la.ka.'pia̯s.to], [la.ka.pi.'a̯s.to], or [la.ka.'pa̯s.to], respectively. Given that there were three response categorizations, the model likelihood was categorical and used a logit linking function. To simplify model interpretation, we report effects in the probability space. The complete model output in the original format is available in the supplementary materials. The responses were modeled in a simple, intercept-only model, and as a function of the post-consonantal glide ([j], [w]). A “hiatus” response was set as the default and the model estimated intercepts for the “simplification” and “triphthong” responses. Figure 4.1 illustrates the overall posterior probabilities of a given response (panel A), and as a function of the glide (panels B and C).

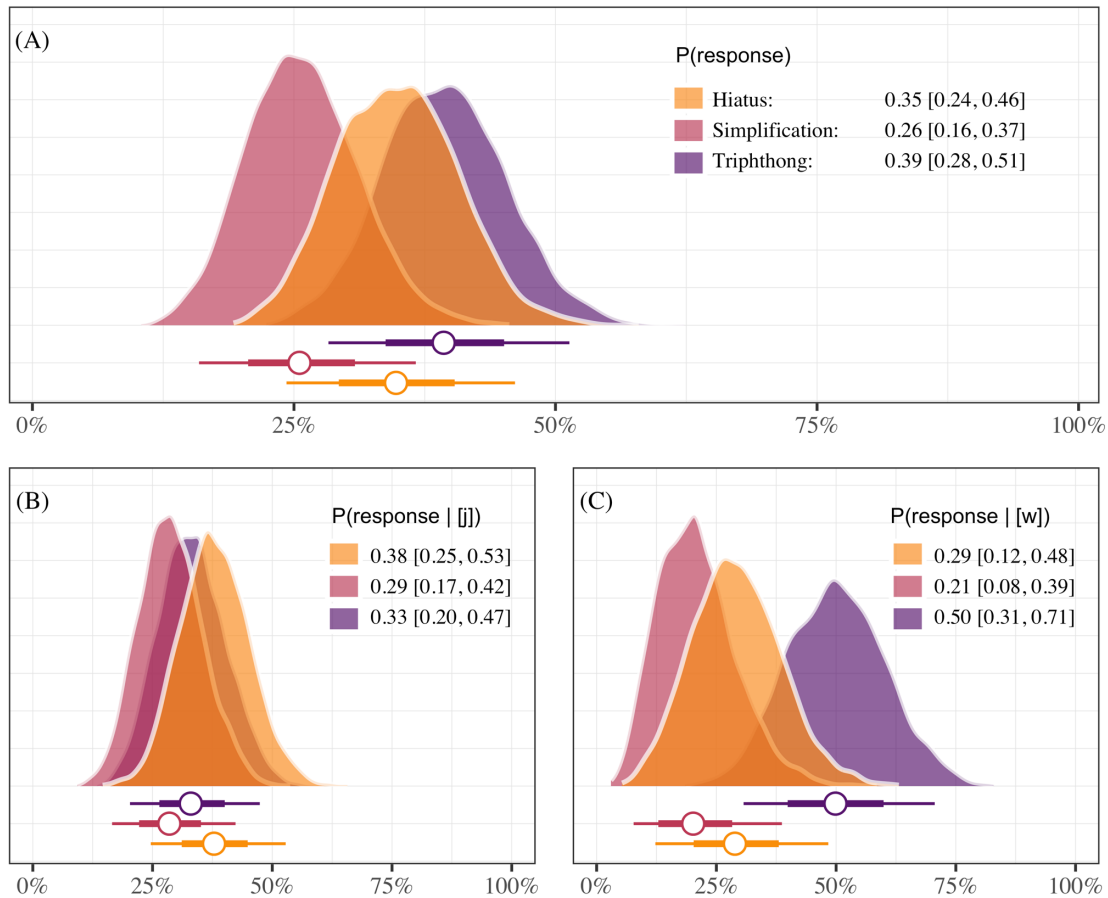


Figure 4.1: Posterior probabilities of responding “hiatus”, “triphthong”, or “simplification”. Each panel plots the posterior medians $\pm 66\%$ and 95% credible intervals. Panel A shows overall responses, and panels B and C show responses as a function of glide type.

With this model we aimed to shed light on how participants responded. Specifically, we were interested in knowing if GVG productions were possible and whether responses depended on the type of glide. We show that GVG responses are indeed possible. Triphthong realizations occurred approximately 39% of the time (0.39 [0.28, 0.51]) in the data set. A production containing a hiatus made up roughly 35% of the data (0.35 [0.24,

0.46]), followed by a simplification of some sort (0.26 [0.16, 0.37]). If we consider realizations as a function of glide type we see that triphthong realizations were possible in both environments, with more being produced with [w] (0.50 [0.31, 0.71]) than with [j] (0.33 [0.20, 0.47]). Overall, the task provides evidence supporting the hypothesis that pre-vocalic glides can be part of the onset in this variety of Spanish because the participants produced triphthongs at least some of the time.

4.4.2 Phrase reading task

For the phrase reading task data we fit models analyzing acoustic properties of the pre-vocalic glide [j]. Importantly, we expect co-occurrence restrictions such that pre-vocalic [j] will be disallowed if preceded by a palatal consonant. In the case that pre-vocalic [j] is indeed blocked after a palatal consonant (i.e., “lliape”), we may observe differences in overall duration of the segment, as well as formant trajectory differences related to height (F1) when compared with a pre-vocalic [j] that is not preceded by a palatal segment (i.e., “piano”). The justification for analyzing duration is due to the prediction that co-occurrence restrictions should disallow contiguous palatal consonant + homorganic glide segments, possibly resulting in an elision. Formant trajectory differences related to height (F1) are expected because in one case (i.e., “piano”) a glide should be produced without restrictions and in the other case (i.e., “lliape”) it should not. That is, in the non-palatal pre-glide items we expect formant movement from the high vocoid [j] to the low vowel [a]. If the pre-vocalic glide is blocked due to a palatal onset then we should not observe the same formant movement.

4.4.2.1 Duration

The duration data were modeled using Bayesian hierarchical linear regression. The model analyzed participant-normalized segment duration as a function of whether the onset included a palatal consonant or not—henceforth *palatal*—, which was sum-to-zero coded (1, -1, labelled as “palatal”, “other”). The model included by-participant and by-item grouping variables, and the *palatal* predictor varied for participants. The model estimates suggest that the presence of a palatal onset is associated with an increase in glide duration ($\beta = 0.22 [-0.20, 0.62]$; MPE = 0.86), though the effect is small and the credible intervals of the posterior distribution are rather wide. Concretely, based on the model, the data, and our prior assumptions, we can conclude with 95% certainty that the effect falls somewhere between -0.20 and 0.62, with an 86% chance that the effect is positive. Figure 4.2 plots the posterior distributions of pre-vocalic glides following [j] (and [w] for comparison) and Figure 4.3 provides a forest plot of the parameter estimates, including all grouping variables. A traditional table summary of the model output is available in the supplementary materials:

https://osf.io/fyt4d/?view_only=1e5f867b1896417cbaa21a2872221cf8.

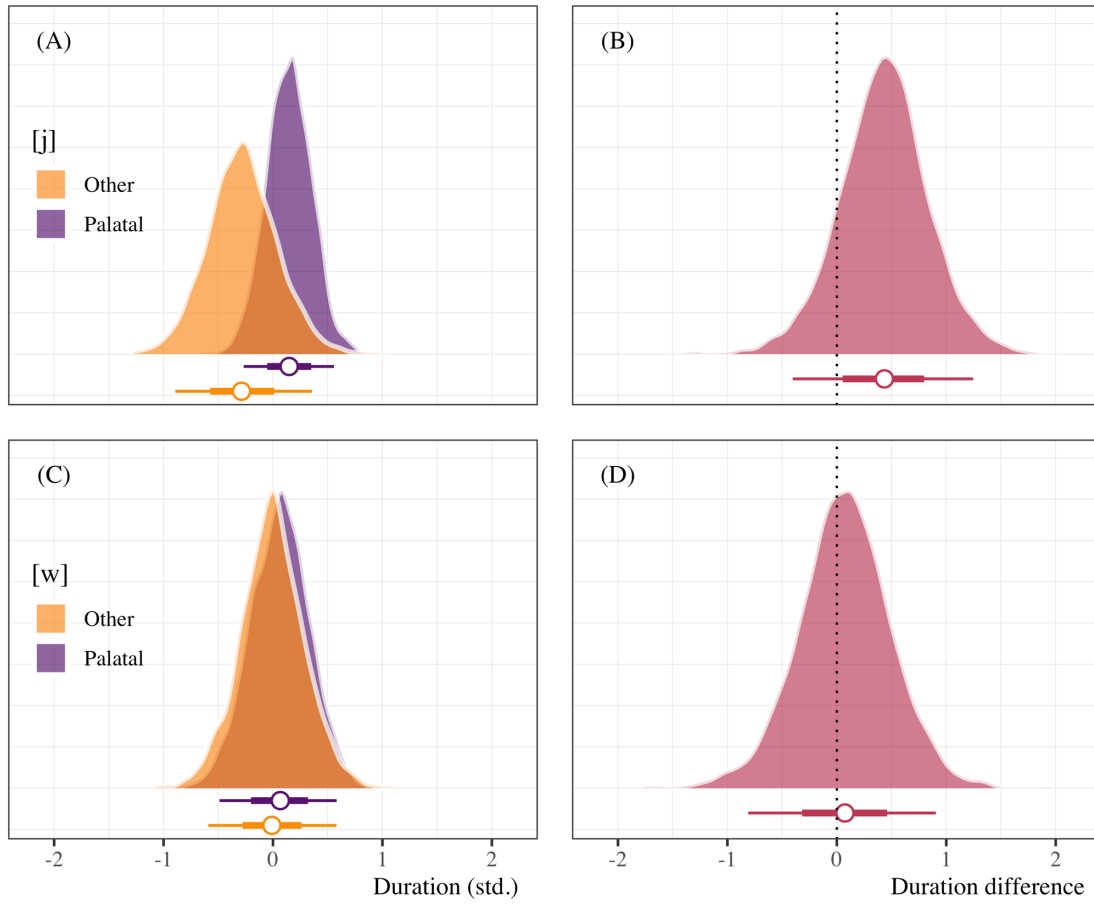


Figure 4.2: Posterior distributions of pre-vocalic glide duration (panels A and C) following palatal and non-palatal onsets, as well as duration difference plots (panels B and D). Points represent posterior medians $\pm 66\%$ and 95% credible intervals.

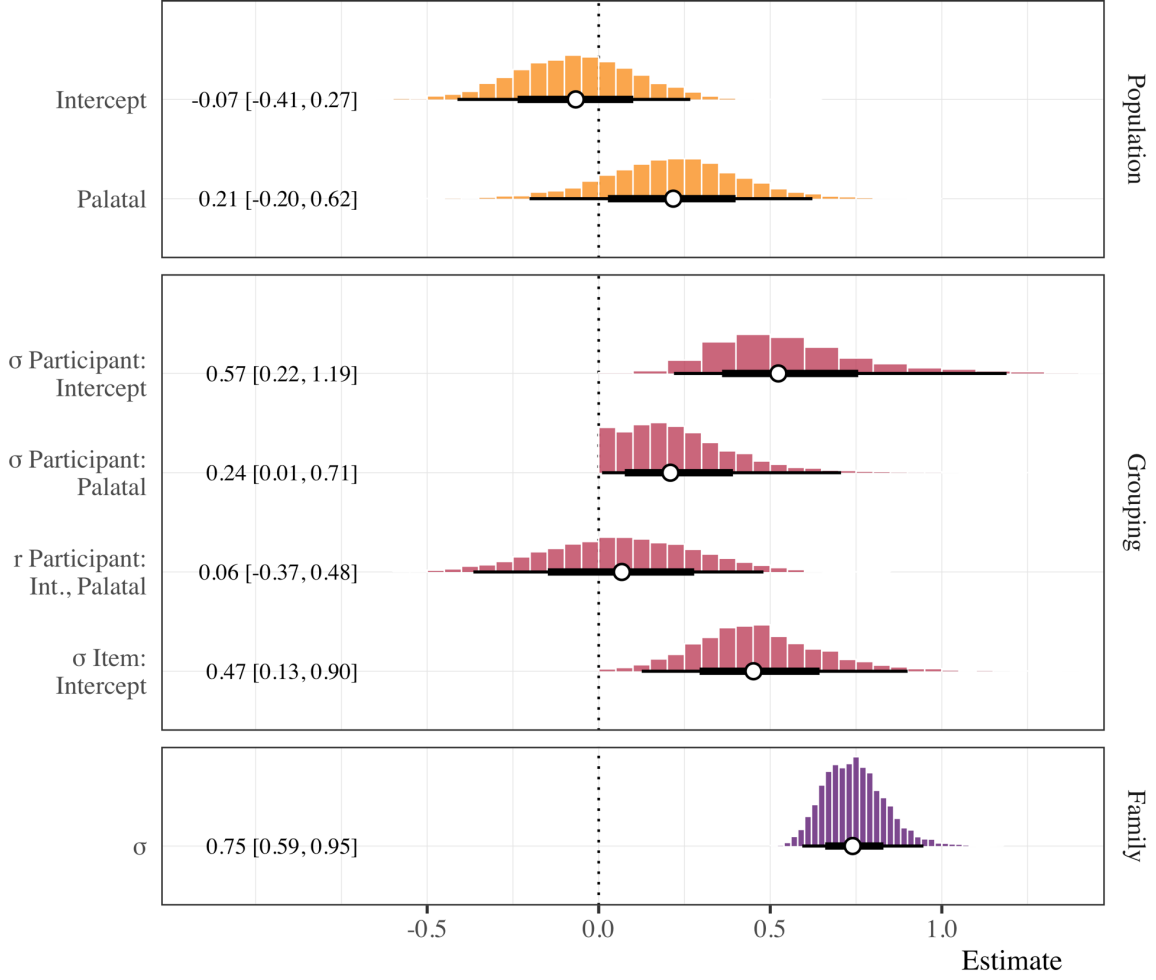


Figure 4.3: Forest plot of posterior distributions of duration model estimates. Points represent posterior medians $\pm 66\%$ and 95% credible intervals.

4.4.2.2 Formant trajectories

The F1 trajectory data were analyzed using a Bayesian Generalized Additive Mixed Model (GAMM, Sóskuthy, 2017; Winter & Wieling, 2016; Wood, 2006). We modeled the participant-normalized F1 values as a function of pre-glide onset, *preceding consonant* (“palatal”, “other”), and a non-linear function of the formant trajectory for the [j] productions. The *preceding consonant* parametric term was set as an ordered variable with *palatal* coded as 0. Cubic regression splines with three basis knots were applied (a) as a

reference smooth to the time course, (b) as a difference smooth to the time course conditioned on the preceding consonant, and (c) as a random smooth for each participant conditioned on the time course. This specification uses the trajectory of the *palatal* condition as the baseline and compares it to the *other* trajectory.

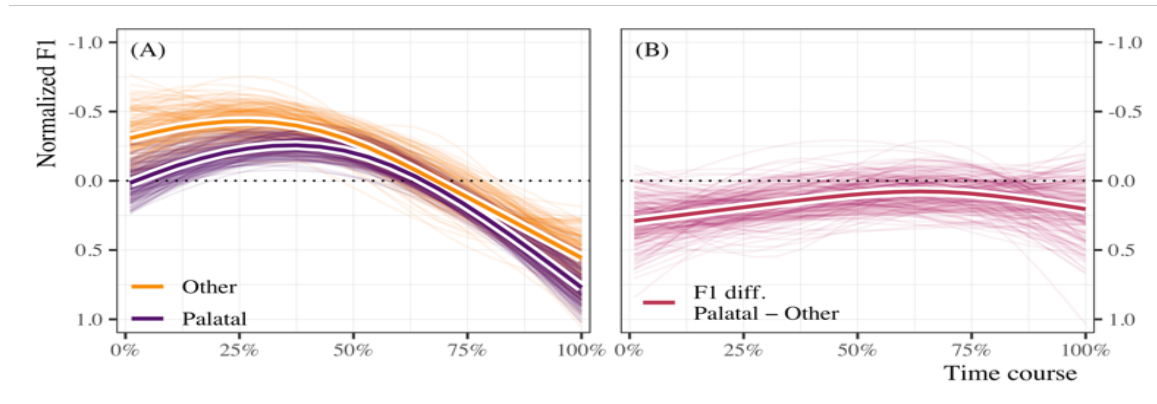


Figure 4.4: Non-linear formant (F1) trajectories of [j] when preceded by palatal and non-palatal onsets (panel A), and estimated differences (palatal – other) in standardized F1 over the time course (panel B). In both panels, the thin lines represent 300 draws of plausible lines from the posterior distribution. The thicker lines outlined in white represent the model average for the population estimate.

Panel A of Figure 4.4 plots 300 posterior draws of plausible lines for the time course of the F1 trajectories in the palatal and other conditions. The thicker lines outlined in white represent the most plausible trajectories based on the data, the model, and our prior assumptions. One observes non-linear trajectories that are generally non-overlapping at the beginning of the time course and tend to converge on a single trajectory after approximately 50% of the time course. The estimates for the difference smooth suggest a small, negative

effect with a moderate amount of uncertainty around the estimate ($\beta = -0.15 [-0.31, 0.00]$; MPE = 0.98). The probability that the effect is negative is approximately .98. The uncertainty around the estimate is manifested in panels A and B of Figure 4.4 by the overlapping colored lines. Panel B plots the estimated difference between the palatal and other conditions. The figure corroborates the estimates from the GAMM, as there appears to be a non-zero difference between trajectories during the first half of the time course, though some plausible estimates indicate that the difference may also be zero. Figure 4.5 provides a forest plot of the model summary. A traditional table summary is provided in the supplementary materials:

https://osf.io/fyt4d/?view_only=1e5f867b1896417cbaa21a2872221cf8.

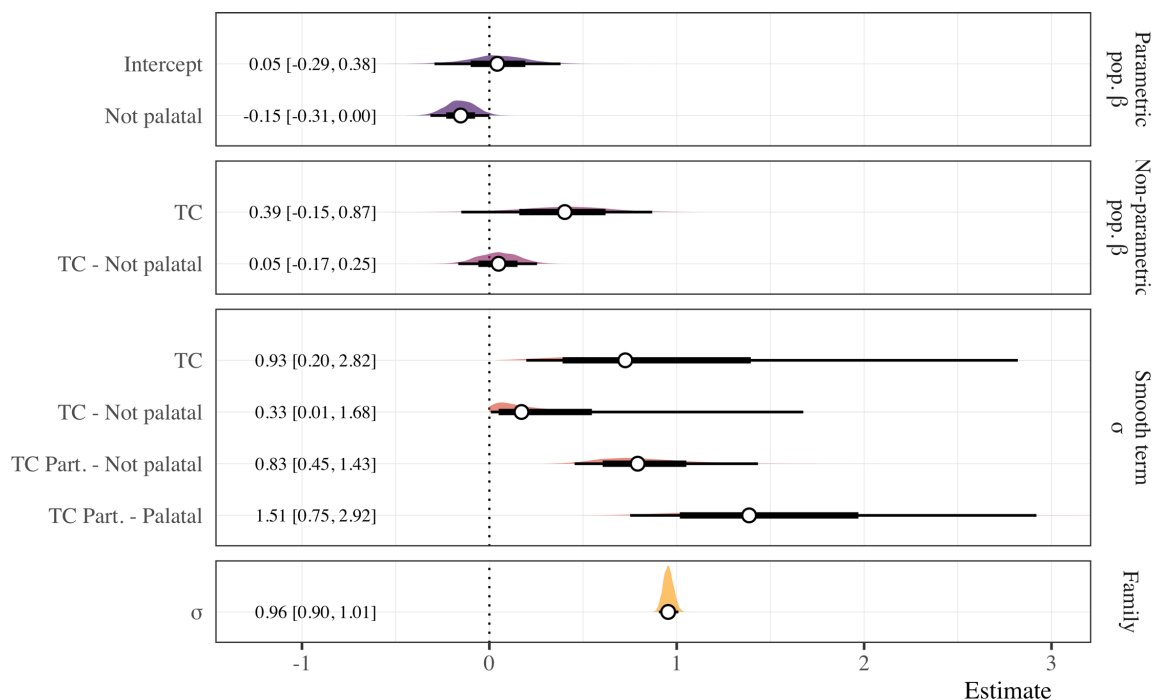


Figure 4.5: Forest plot of posterior distributions from the GAMM analyses of the F1 time course for [j]. Points represent posterior medians $\pm 66\%$ and 95% credible intervals for

which numeric summaries are provided in the left margin. Vertical faceting distinguishes between parametric (top) and non-parametric population estimates, followed by grouping estimates for smooth terms and overall standard deviation.

5. Discussion

The results of the production experiments suggest that Sonoran speakers used variable strategies when producing the CGVGC sequences. Importantly, they produced the GVG sequence at least some of the time in the syllabification task, indicating that this realization is indeed possible, though not categorical. As noted with regard to the syllable division task (4.4.1), the labiovelar glide is more commonly pronounced in a triphthong than its coronal counterpart, which could be a function of the more frequent labial and labiovelar consonants in the stimuli. An acoustic analysis of the production data from the phrase reading task indicates that pre-vocalic [j] may be realized with subtle differences after palatal consonants. Taken together, the duration and formant trajectory analyses suggest (1) that participants may be producing slightly *longer* pre-vocalic [j] when preceded by a +palatal onset segment, and (2) that the starting point for the F1 transition from [j] to [a] is slightly lower when the onset contains a +palatal segment. Concretely, [j] was not shortened or elided as hypothesized due to co-occurrence restrictions. On the contrary, we find evidence that the hypothesized co-occurrence restriction may result in lengthening of the onset segment. One possibility is that the participants may be producing a slightly longer segment, as opposed to eliding it (making it shorter), in the palatal condition because of the fact that they cannot naturally produce both. That is, our exploratory analysis of the

[j] time course suggests that the tendency to increase duration could be a strategy to avoid the illicit sequence.

In sum, the research question of this study can be answered in the affirmative: a postconsonantal, prevocalic glide can be parsed as the second segment of a complex onset in Sonoran Spanish, at least some of the time, demonstrating dialect-internal, as well as possibly, cross-dialectal variation. This finding supports the position (presented in Section 2) that the debate in the literature over syllabic affiliation of on-glides has been oversimplified by trying to decide in favor of either an onset or a nuclear parse for Spanish postvocalic glides, rather than allowing for both possibilities under different conditions. Future studies need to investigate the conditions which favor onset parsing over simplification or hiatus, on the one hand, and over nuclear parsing on the other and whether the nature of the task could influence the results (perhaps making the lexical and postlexical distinction relevant). Similar studies on other dialects could also shed additional light on the behavior of onglides.

The variation found in this study can be easily captured by an optimality-theoretic account.

(3) DEP, IDENT[son] >> *S-HeavyRH >> ONSET, MAX-IO >> *ONSET/glide, *NUC/glide
Constraints such as *ONSET/glide and *NUC/glide penalize the association of glides to the onset and nucleus positions, respectively. The unranked status of these two constraints with respect to each other predict the possibility of parsing a glide either in the onset or the nucleus (3), dependent upon other factors (and higher ranked constraints), such as the availability of an onset or the size of the rhyme, captured by the constraints are ONSET and *S-HeavyRH. ONSET requires that all syllables have an onset. *S-HeavyRH is shorthand

for the ban on superheavy rhymes of more than three constituents, i.e. rhymes cannot have two complex subconstituents, or in other words, a complex nucleus followed by a complex coda is not possible, only the nucleus or the coda can be complex. MAX-IO_μ demands that input moras be preserved in the output and MAX-IO penalizes deletion

laca/piaisto/ [pjájsto]

	*S-HeavyRH	ONSET	MAX-IO	*ONSET/glide	*NUC/glide	MAX-IO _μ
a. <i>☞</i> pjájs.to				*		**
b. pi.ájs.to		*!				*
c. pjás.to			*!			
d. pjájs.to	*!				*	**

Figure 5.1. Optimality-theoretic selection of a candidate with postconsonantal onset glide

Figure 5.1 shows how the constraint ranking in (3) selects a candidate with an onset glide for one of the nonce words used in the experiment. Despite a violation of *ONSET/glide, candidate (a) is the winner because (b), (c), and (d) violate higher ranked constraints, ONSET, MAX-IO and *S-HeavyRH, due to lack of an onset, deletion, and a superheavy rhyme (with a complex nucleus and a complex coda), respectively.

As indicated in the findings, but not shown in the visual representation provided by the tableaux, the selection of (a) is not categorical. (b) and (c) are also possible in the syllabification task. This variability can be modelled in Stochastic OT (Boersma and Hayes 2001). Stochastic OT views constraint rankings as ranges of values, which can overlap and result in a ranking reversal, when a selection point of the higher ranked constraint falls in the lower end of the range and that of the lower ranked constraint is in the higher end of

the range (Boersma and Hayes 2001). Variation in the parsing of glides as onsets, glide deletion or hiatus in Sonoran Spanish points to overlapping rankings of the constraints ONSET, MAX-IO, and *ONSET/glide.

Dialects that, unlike Sonoran Spanish, do not allow onset glides are predicted to rank *ONSET/glide higher than ONSET and/or MAX-IO and would have hiatus or deletion instead of triphthongs (with the glide parsed in the onset) (4). Whether there is hiatus or deletion needs to be confirmed empirically. Although the agreed-upon knowledge is that these varieties do not allow glides as the second member of an onset, it remains to be determined if a prevocalic glide could be parsed as such as an alternative to a superheavy rhyme. As mentioned above, glides in intervocalic or syllable initial position become consonantal (i.e., onset strengthening) in these dialects. This reflects the ranking of IDENT[son] below *ONSET/glide and above *NUC/glide (5). A consonantized glide ([-son]) violates IDENT[son], which bans changing the [sonorant] specification of the input.

(4) *S-HeavyNUC >> *ONSET/glide >> ONSET, MAX-IO >> *NUC/glide

(5) *ONSET/glide >> ONSET, MAX-IO >> IDENT[son] >> *NUC/glide

Finally, a few words about study limitations are in order. Although the small number of participants must be acknowledged as a limitation of this study, it is likely that a larger sample would not alter the finding that postconsonantal glides can be parsed in the onset in Sonoran Spanish. Further research is needed to determine whether variation regarding the affiliation of postconsonantal prevocalic glides can also be found in other dialects. This study should be replicated with additional subjects (controlled for gender, age, linguistic

profile), and, for comparative purposes, with speakers of dialects that do not allow onset glides.

6. Conclusion

This study examined the syllabic affiliation of prevocalic glides in Sonoran Spanish. Since glides are licit in the onset (i.e., no onset strengthening in Sonoran Spanish), a postconsonantal, prevocalic glide could be potentially parsed as the second segment of a complex onset in this variety. Experimental evidence showed that Sonoran speakers produced the GVG sequence at least some of the time and thus it is possible for it to be parsed as the second segment in a complex onset. Furthermore, acoustic analyses demonstrated a tendency for pre-vocalic segments to be longer after palatal consonants and the duration increase was shown to be due to lengthening of the onset (a strategy to avoid an illicit sequence). Glides can be *variably* parsed in a complex onset. We conclude that the affiliation debate has been oversimplified by considering either an onset or a nuclear parse.

Additionally, onset glides in Sonoran Spanish provide evidence for cross-dialectal and intra-dialectal variation in syllabic affiliation. It is shown that an optimality theoretic model can capture and explain the sources of internal variation as competing factors which result in variation in output parsing.

Finally, we gained insight on the status of glides in this variety of Spanish and in Spanish phonology in general. Additionally the findings have implications for research on monolingual and bilingual acquisition and on acquisition errors (e.g., language learning,

speech pathology, etc.), as work in these areas is based on phonological descriptions that could be inaccurate or inapplicable to all varieties.

Appendix

Syllabic affiliation of prevocalic (postconsonantal) glides in Sonoran Spanish

Hypothesis 1

C+G+V+G	Nonce words	C+G(round)+V+G	
pjaj	lacap[jaj]sto	pwaj ¹ (if onset restrictions, it should not be acceptable)	lacap[waj]sto
bjaj	lacab[jaj]sto	bwaj	lacab[waj]sto
fjaj	lacaf[jaj]sto	fwaj	lacaf[waj]sto
tjaj	lacad[jaj]sto	twaj	lacad[waj]sto
djaj	laced[jaj]sto	dwaj	laced[waj]sto
kjaj	lacak[jaj]sto	kwaj	lacak[waj]sto
gjaj	lacag[jaj]sto	gwaj	lacag[waj]sto

Hypothesis 2

nonce words

[tʃ]__

ch[j]eba

meboch[j]ena

[j]__

ll[j]ape

¹ If there are coarticulation restrictions in the onset, pwaj, bwaj, fwaj should not be acceptable, but the rest (twaj, dwaj, kwaj, gwaj) should be fine

ruy[j]ola

pay[j]elo

costoñ[j]elo

[n]__

mañ[j]ela

References

- Alvar, Manuel. 1996. *Manual de dialectología hispánica*. Barcelona: Ariel.
- Anderson, Raquel T. 2002. Onset clusters and the sonority sequencing principle in Spanish: A treatment efficacy study. In Fay Windsor, M. Louise Kelley and Nigel Hewlett (eds.), *Investigations in clinical phonetics and linguistics*, 213- 224. Mahwah, NJ: Lawrence Erlbaum Associates.
- Baković, Eric. 2006. Hiatus resolution and incomplete identity. In Fernando Martínez-Gil and Sonia Colina (eds.) *Optimality-theoretic studies in Spanish phonology*, 62-73. Amsterdam and Philadelphia: John Benjamins.
- Barlow, Jessica. 2005. Phonological change and the representation of consonant clusters in Spanish: A case study. *Clinical Linguistics and Phonetics* 19. 659–679.
- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.
- Boersma, Paul and Bruce Hayes 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45-86.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28. <https://doi.org/10.18637/jss.v080.i01>

Canfield, D. Lincoln. 1981. *Spanish Pronunciation in the Americas*. Chicago: U. of Chicago P.

Clements, George N., and Keyser, Samuel J. 1983. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Colina, Sonia. 2009. *Spanish Phonology: A syllabic perspective*. Washington, DC: GUP.

Colina, Sonia 2012. Syllable structure. In Hualde, José I., Antonio Olarrea and Erin O'Rourke, (eds.), *Handbook of Hispanic linguistics*. Oxford, UK: Blackwell. 133-151.

Colina, Sonia and Miquel Simonet. 2014. Galician coda restrictions and plural clusters. *Linguistics* 52(6). 1433-1460.

Gelman, Andrew, Daniel Simpson, and Michael Betancourt. 2017. The prior can often only be understood in the context of the likelihood. *Entropy* 19(10). 1–13. <https://doi.org/10.3390/e19100555>

Harris, James W. 1983. *Syllable structure and stress in Spanish*. Cambridge, MA: MIT Press.

Hualde, José I. 1991. On Spanish syllabification. In Héctor Campos and Fernando Martínez-Gil (eds.), *Current studies in Spanish linguistics*, 475-493. Washington, DC: Georgetown University Press.

Hualde, José I. 1997. Spanish /i/ and related sounds: an exercise in phonemic analysis. *Studies in the Linguistic Sciences* 27. 61-79.

Hualde, José I. 1999. Patterns in the lexicon: hiatus with high unstressed high vowels in Spanish. In Javier Gutiérrez-Rexach and Fernando Martínez-Gil (eds.), *Advances in Hispanic linguistics. Papers from the 2nd Hispanic Linguistics Symposium*, 182-197. Somerville, MA: Cascadilla.

Hualde, José I. 2002. On the diphthong/hiatus contrast in Spanish: some experimental results. *Linguistics* 40. 217-234.

Hualde, José I. 2005. *The Sounds of Spanish*. Cambridge, UK: CUP.

Hualde, José I. 2014. La silabificación en español. In Rafael Núñez-Cedeño, Sonia Colina, and Travis Bradley (eds.), *Fonología generativa contemporánea de la lengua española*, 195-215. Washington, DC: Georgetown UP.

Hutchinson, Sandra P. 1974. Spanish vowel sandhi. *Papers from the parasession on natural phonology*, 184-192. Chicago: Chicago Linguistic Society.

- Kehoe, Margaret, Geraldine Hilaire-Debove, Katherine Demuth and Conxita Lleó. 2008. The structure of branching onsets and rising diphthongs: Evidence from the acquisition of French and Spanish. *Language Acquisition* 15. 5-57.
- Martínez-Gil, F. 2000. La estructura prosódica y la especificación vocálica en español: el problema de la sinalefa en ciertas variedades de la lengua coloquial contemporánea. In Juana Gil Fernández (ed.), *Panorama de la fonología española actual*, 511-560. Madrid: Arco Libros, S.L
- Martínez-Gil, Fernando. 2016. Syllable merger in Chicano Spanish: a constraint-based analysis. In Rafael Núñez-Cedeño (ed.), *The Syllable and Stress: Studies in Honor of James W. Harris*, 139-181. Boston: Mouton de Gruyter.
- Morales-Front, Alfonso I. 2014. El acento. In Rafael Núñez-Cedeño, Sonia Colina, and Travis Bradley (eds.), *Fonología generativa contemporánea de la lengua española*, 235-265. Washington, DC: Georgetown UP.
- Piñeros, Carlos-Eduardo. 1998. Prosodic morphology in Spanish: Constraint interaction in word-formation. Columbus, OH: The Ohio State University dissertation.
- Prieto, Pilar. 1992. Morphophonology of the Spanish diminutive formation: A case for prosodic sensitivity. *Hispanic Linguistics* 5. 169-205.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Shelton, Michael. 2007. *An experimental approach to syllable weight and stress in Spanish*. College Station, PA: The Pennsylvania State University dissertation.

Shelton, Michael, Chip Gerfen and Nicolás Gutiérrez Palma. 2010. How Spanish phonotactics informs psycholinguistic models of speech production. In Karlos Arregi, Zsuzsanna Fagyal, Silvina A. Montrul and Annie Tremblay (eds.), *Romance linguistics 2008: Interactions in Romance*, 249-264. Amsterdam: John Benjamins.

Shelton, Michael, Chip Gerfen and Nicolás Gutiérrez Palma. 2012. The interaction of subsyllabic encoding and stress assignment: A new examination of an old problem in Spanish. *Language and Cognitive Processes* 27(10). 1459-1478, DOI: [10.1080/01690965.2011.610595](https://doi.org/10.1080/01690965.2011.610595)

Sóskuthy, Márton. 2017. *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction*. arXiv Preprint arXiv:1703.05339.

Stan Development Team. 2018. Stan modeling language users guide and reference manual (Version 2.18.0). Stan Development Team. Retrieved from <http://mc-stan.org>

Winter, Bodo, and Martijn Wieling. 2016. How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *Journal of Language Evolution* 1(1). 7–18.

Wood, Simon. 2006. *Generalized additive models: An introduction with R* (2nd ed.). Boca Raton: CRC Press.