

Article

Spanish in the Southeast: What a Swarm of Variables Can Tell Us about a Newly Forming Bilingual Community

Jim Michnowicz *, Rebecca Ronquest, Sarah Chetty, Georgia Green and Stephanie Oliver

Department of World Languages and Cultures, North Carolina State University, Raleigh, NC 27695-8106, USA

* Correspondence: michnowicz@ncsu.edu

Abstract: The southeastern United States has experienced rapid growth in the Hispanic population in recent decades, giving rise to a newly forming bilingual community. The present study builds on previous work by the authors via expansion of a “variable swarm”: the analysis of multiple linguistic variables simultaneously for the same set of speakers, with the goal of understanding patterns of accommodation and change within the community. The initial study included four linguistic variables (prosodic rhythm, bilingual discourse markers, the realization of /bdg/ and vowel space), and the present study adds an additional four variables (bilingual filled pauses, subject pronoun realization, code switching, and the labiodental realization of orthographic <v>) for 23 speakers of Mexican and Central American origin across two sociolinguistic generations (G1 vs. G2). Results for individual speakers show a pattern of adoption of some features by speakers of both generations (such as English-influenced prosodic rhythm and phonological filled pauses), while other, possibly more salient forms directly integrated from English (English discourse markers and code switching) exhibit later, highly variable rates of adoption, suggesting that speakers may consciously manipulate these variables as part of a process of active identity construction. Likewise, G1 speakers show fewer correlations among linguistic variables than G2 speakers, and patterns reveal that some bilingual forms are incorporated in tandem due to shared phonological traits or discourse functions. The innovative swarm analysis further contributes to the advancement of techniques employed in sociolinguistic research by serving as a bridge between traditional first- and second-wave studies that focus on a single variable, and third-wave studies that focus more on variation at the individual level.



Citation: Michnowicz, Jim, Rebecca Ronquest, Sarah Chetty, Georgia Green, and Stephanie Oliver. 2023. Spanish in the Southeast: What a Swarm of Variables Can Tell Us about a Newly Forming Bilingual Community. *Languages* 8: 168. <https://doi.org/10.3390/languages8030168>

Academic Editors: Whitney Chappell and Sonia Barnes

Received: 9 December 2022

Revised: 21 June 2023

Accepted: 29 June 2023

Published: 14 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Spanish in the U.S.; sociolinguistics; language contact; “New Destination” communities; “variable swarm”; interfacing variables

1. Introduction

The southeastern United States, and North Carolina in particular, is home to one of the most rapidly growing Hispanic populations in the country, having witnessed 900% growth between 1990 and 2010 (Carolina Demography 2021). Due to a booming economy among other factors, as of 2020, North Carolina has an estimated Hispanic population of over one million (Pew Research Center 2014). Unlike other regions within the U.S., such as the Southwest and Northeast, which have long-standing, well-established communities, the Southeast represents a newer community or “New Destination” community (Zúñiga and Hernández-León 2005), in which the effects of language and dialect contact are still taking shape. Such a community provides the opportunity to examine the initial stages of language and dialect contact as they emerge in real time. The present study aims to augment our understanding of the diverse and rapidly growing Hispanic communities in the southeastern United States, with particular emphasis on North Carolina, by extending our initial study (Ronquest et al. 2020) via the analysis of four additional linguistic variables in a “variable swarm” (Thomas 2015): subject pronoun expression (SPE), code switching, phonological filled pauses, and an acoustic analysis of the pronunciation of and <v>.

As discussed in the original “variable swarm” analysis, many sociolinguistic studies focus on a single linguistic variable and assess the relationships between the variable and macro-sociolinguistic factors. A swarm, in contrast, analyzes multiple variables for the same set of speakers and is able to provide a more detailed and nuanced view of the speech of a community, as well as offer insight into particular patterns of speaker variation and how the variables intersect and interact. Thomas (2015, p. 3) states succinctly that “[e]ven if one variable shows noteworthy patterns that provide clues about social identities of speakers, it cannot provide a complete picture of the intersecting identities that individuals exhibit. Each linguistic variable may reveal new social meanings and patterning. What is needed is inquiry that compares a large number of diverse variables”.

This study’s methodology and general line of inquiry, which focuses on how multiple variables pattern across and within individual speakers and larger social groups (in particular, sociolinguistic generation), situates it at the border of second- and third-wave sociolinguistics. Eckert, in her groundbreaking 2012 article, identifies three “waves” of variationist sociolinguistics, with the second and third waves overlapping in time and currently constituting the bulk of variationist work. According to Eckert’s (2012) definition, first-wave studies of the 1960s and 1970s focused on how linguistic variables index pre-determined socioeconomic groups, with variation “resulting from the effects of these categories on speakers’ orientation to their assigned place in the hierarchy” (p. 90). Stylistic variation was primarily seen as the avoidance of stigmatized forms in particular contexts. Second-wave studies, as defined by Eckert (2012), returned to the ethnographic roots of the very first variationist study (Labov 1963), with a focus on how speakers within communities can actively utilize “the vernacular as an expression of local or class identity” (p. 91), and language is but one (albeit key) component in the construction of group identity (along with choices in clothing and music, for example). Finally, third-wave studies shift the focus from group behavior to individual speaker behavior, exploring how “speakers place themselves in the social landscape through stylistic practice” (Eckert 2012, p. 94), a view which emphasizes the agency that speakers have in determining not only their own systems of linguistic variation, but ultimately those of the speech communities to which they belong. The present study serves as a bridge between the (sometimes) artificial distinctions between “waves.” By means of an examination of eight sociolinguistic variables through a largely second-wave methodological and theoretical lens, we are able to begin to construct a larger picture of how discrete linguistic variables from across multiple domains of language (segmental and suprasegmental phonology/phonetics, morpho-syntax, and pragmatics) pattern across and within individual speakers. The swarm approach allows us to observe which variables are adopted, either in conjunction or in isolation, by particular speakers, illuminating both individual variation as well as each speaker’s linguistic behavior as part of a larger social group (in this case, primarily sociolinguistic generation). In this way, a swarm approach to language variation provides a panoramic view of a speech community that is not possible through the analysis of only one or two variables, as well as sets the stage for future third-wave studies that can further explore why speakers adopt or reject possible contact forms in a newly developing bilingual region.

We begin with a brief description of Spanish in the southeastern U.S. and summarize several of the key studies that have been conducted in the area to date. Next, the overall methodology (e.g., speakers and corpus) is described, followed by separate subsections for each of the four variables under investigation. The results of the present four variables are then combined with the initial four variables from Ronquest et al. (2020), and an exploration of the interaction of all eight variables is presented in Section 4. The paper ends with a general discussion, directions for future research, and conclusions.

2. Background Studies

2.1. The Southeast as a New Dialect Region/“New Destination” Community

Since the late 1990s, the Hispanic population in the southeastern United States has grown significantly. According to the latest data, North Carolina is one of four U.S. states

that has experienced an increase to over 1 million Hispanics/Latinos since 2010, with an estimated population of 1,118,596 in 2020 (Carolina Demography 2021). Just over half (56%) of Hispanics residing in NC were born in the U.S., and 44% are foreign born. While 55% are of Mexican origin, the demographic profile is diversifying. Central American speakers make up the second-largest group, especially those of El Salvadoran and Guatemalan descent (16%) (U.S. Census Bureau 2020).

Scholars investigating Spanish in the United States have tended to focus their attention on regions within the country that have well-established communities such as the upper-Midwest, Southwest, and Northeast (c.f. Otheguy and Zentella 2012; Poplack 1978; Silva-Corvalán 1994; among others). Within the past decade in particular, however, the rapidly growing Hispanic/Latino population in the Southeast has motivated research in this previously understudied region, encompassing a wide range of topics. Studies of rhythmic differences (Carter 2005; Ronquest et al. 2020), including Hispanic English (Wolfram et al. 2004, 2011), have revealed distinctions in rhythmic profiles among bilingual and monolingual speakers of distinct backgrounds. Limerick's (2019, 2021) investigations of subject pronoun expression (SPE), which will also be examined in the present study and are described in more detail below, have revealed that Hispanics residing in Georgia exhibit distinct patterns of usage from those residing in other U.S. regions, indicating that their system is at a different/intermediate stage of development than those in more established communities.

In the lexical domain, Michnowicz et al. (2018) reported differences in acceptance and usage of English-origin loan words among first- and second-generation speakers of varying backgrounds residing in North Carolina, as well as differences with other varieties of Spanish in the U.S. Finally, work on language attitudes (Montes-Alcalá and Sweetnich 2014; Howe and Limerick 2020; Knouse et al. 2022), identity (Carter 2007, 2013), attitudes towards inclusive language (Michnowicz et al. 2023a), and language maintenance and shift (Michnowicz et al. 2023b) suggests that a complex interplay of factors is shaping the way in which residents in the Southeast perceive and utilize their language(s), as well as strategies they employ to construct their identity in a nascent bilingual environment. In conjunction, the investigations conducted thus far confirm that the Southeast is home to diverse linguistic communities at various stages of development, and that multiple factors—linguistic, social, and attitudinal—are involved in the creation of novel contact varieties.

2.2. *New Dialect Formation in Language Contact Settings*

Spanish in the United States is subject to two related but distinct forces that can shape the direction of future development of the language varieties present within a community (Otheguy et al. 2007). The first of these is contact with English, as bilingual speakers adopt English-influenced forms from the majority/dominant language, not just as a way to index a bilingual identity (Zentella 1997), but also as a strategy of “lightening the cognitive load of having to remember and use two different linguistic systems” (Silva-Corvalán 1994, p. 6). Language contact often results in convergence, where bilingual speakers produce forms that are in some way intermediate between the two languages in contact, particularly at the levels of pragmatics and lexicon (Silva-Corvalán 2008) or phonology (Ronquest 2012; Carter and Woldford 2016).

The second force that plays a role in the development of Spanish in the U.S. is dialect contact with other varieties of Spanish. Dialect contact between mutually intelligible linguistic varieties can result in linguistic accommodation, as speakers negotiate differences in form across regional dialects (Britain and Trudgill 1999). Accommodation is often, but not exclusively, in the direction of the majority or prestige dialect in the community, as has been found for pronouns of address (Hernández 2002) and subject pronouns (Otheguy et al. 2007). Dialect contact often results in processes of leveling and koineization or new dialect creation, whereby after a period of heightened variation, differences between dialects are diminished across time, as the younger generations in a community converge on a new set of dialect norms (Kerswill 2013). Koineization involves the processes of mixing, leveling,

and simplification, which can result in “the reduction or attrition of marked variants” (Trudgill 1986, p. 98).

Spanish in the U.S. presents a mixture of the two processes of language and dialect contact, and the line between them can often blur. For example, in her study of Spanish in New York, Zentella (1990) found that speakers of different Spanish varieties often opted for an English loanword in order to facilitate communication across communities. In other words, speakers looked to a borrowed form in order to resolve a difficulty arising from dialect contact. Particularly with regard to Spanish in the U.S., where speakers are exposed to both English and different varieties of Spanish, language contact and dialect contact are two sides of the same coin, since the consensus or intermediate form may well involve an English loan, rather than a borrowing from one of the Spanish dialects in contact. The entire sociolinguistic context must be taken into account, as speakers negotiate and accommodate both towards English and towards other varieties of Spanish simultaneously.

Features such as markedness or salience of particular forms play an important role in which features are adopted in a new (bilingual) community. Trudgill (1986) states that “[i]n contact with speakers of other language varieties, speakers modify those features of their own varieties of which they are most aware” (p. 11). Two of Trudgill’s (1986, p. 11) criteria for salience of linguistic forms which are particularly relevant to the current study are if a form is “overtly stigmatized” within a community and if two forms are “phonetically radically different.” We will argue later that these criteria apply to language contact-induced forms as well. Erker (2017) explicitly tests this idea, finding dialect convergence for two highly salient variables, /s/ weakening and *voseo*, while differences based on region of origin persisted for a low-salience variable, subject pronoun expression (SPE). Erker (2017) expands on Trudgill’s (1986) salience criteria, finding that the linguistic domain of a feature (e.g., phonology vs. morpho-syntax) is not the deciding factor, but rather “what accounts for the different fate of features in settings characterized by both dialectal and language contact is their varying social salience” (p. 15). As we will see in the results, social salience also appears to play an important role in determining the order in which linguistic features are adopted by the newly forming Spanish-speaking community in NC.

2.3. The Initial Variable Swarm

The initial variable swarm (Ronquest et al. 2020), which serves as the basis for the present study, analyzed the lenition of /bdg/, vowel production (including the size of a speaker’s vowel space as analyzed by the Convex Hull Area—CHA), prosodic rhythm (nPVI), and bilingual discourse markers (DMs) in the speech of the same 23 participants examined herein. Each of these initial variables is detailed briefly in the sections that follow, and we refer the reader to the original study for more detail.

2.3.1. Realization of Intervocalic /bdg/

In many monolingual dialects of Spanish, /bdg/ show two realizations that exist in complementary distribution: stop [bdg] is found after pauses and homorganic consonants, whereas approximant [βðɣ] arise in all other contexts, including in intervocalic position (Hualde 2005). The lenition of /bdg/ > [βðɣ] is a gradient, acoustic phenomenon that responds to both linguistic and social factors, including dialect and whether or not the variety of Spanish is in contact with another language (Colantoni and Marinescu 2010; Hualde et al. 2011; Lipski 1994, 2020).

Mexican and Central American Spanish, the two varieties studied here, both demonstrate stronger, more stop-like realizations of /bdg/ in intervocalic position when compared to some other varieties, such as Caribbean Spanish (Lipski 1994). Regarding Spanish varieties in the United States, the question of English influence arises, since English only has the stop variants of /bdg/. Some research has shown that heritage speakers (second generation [G2] or later) largely match monolingual patterns of /bdg/ lenition (Knightly et al. 2003), while other studies show a great deal of individual variation, based at least in part on how often a G2 speaker uses and interacts with Spanish. Specifically, Rao (2015)

found that regular users of Spanish showed patterns similar to those of monolinguals, while those with less regular exposure showed larger differences, including more stop-like [bdg]. Rao (2015, p. 66) notes that this can contribute to “a potential heritage accent” for some G2 speakers.

The initial swarm analysis (Ronquest et al. 2020) analyzed the intensity difference (IntDiff, Hualde et al. 2011) between the consonant and the following vowel, where a larger intensity difference suggests a stronger, more occlusive-like variant, and a smaller intensity difference indicates a more lenited variant. A total of 15,828 tokens of /bdg/ were analyzed, with results of the mixed-effects linear regression (random intercept of speaker) showing that while there was no significant main effect for any of the social factors (region of origin, generation, sex), there was a significant interaction between consonant (/bdg/) and region/generation. Overall, Mexican speakers showed higher rates of lenition than Central American speakers, but the patterns among G1 and G2 speakers were reversed across dialect groups. Central American G2 speakers produced significantly larger intensity differences than their G1 counterparts, matching the expected result for English-language influence, while for Mexicans the opposite pattern was observed. These results show the complex interplay of factors in determining the outcome of bilingual lects.

2.3.2. Vowel Space and Convex Hull Area (CHA)

The Spanish vowel system, which consists of five phonemes /ieaou/, has traditionally been described as fairly stable across dialect regions (Hualde 2005; Navarro Tomás 1918). The presence of minor differences in vowel quality and quantity across varieties has been established (e.g., Chládková et al. 2011; Quilis and Esgueva 1983), however, not to the extent observed for English, which is characterized by a larger vowel system that is highly variable across geographic regions. Studies of bilingual vowel systems, however, have revealed that both L2 learners of Spanish and G2 speakers of Spanish differ with regard to their pronunciation of the Spanish vowels. The high back vowel /u/ in particular is subject to a more fronted articulation (i.e., higher F2) in both learner and heritage systems in comparison to monolingual norms (Alvord and Rogers 2014; Cobb and Simonet 2015; Menke and Face 2010; Ronquest 2012; Willis 2005; among others). The low mid /a/ has also been described as fronted and approximating /æ/ (Willis 2005), and /e/ is often produced farther back in the vowel space among heritage speakers (Ronquest 2012).

Such differences in vowel quality, acoustic distribution, and organization often result in a more condensed vowel space for L2 learners of Spanish and some bilinguals (Menke and Face 2010), motivating the analysis of the overall area of the vowel space in the first version of the swarm. The Convex Hull Area (CHA), or overall geometric area of the vowel space, was calculated for each speaker. We hypothesized that contact-induced modifications in vowel production—such as centralization or less peripheral point vowels—would result in a smaller CHA, and would most likely be observed among the G2 speakers in the swarm given their greater degree of contact with English. While the mixed-effects linear regression (random intercept of speaker) did not reveal significant differences among the G1 and G2 participants, additional examination of the amount of variation in CHA did indicate more variability among the G2 in comparison to the G1, who were much more consistent in their productions. Analysis of the swarm indicated that 58% (7/12) of G2 and 17% (2/12) of G1 speakers favored a smaller vowel space.

2.3.3. Prosodic Rhythm (nPVI)

One of the well-known differences between Romance languages, such as Spanish, and Germanic languages, such as English, is in prosodic rhythm. In Spanish, a prototypical syllable-timed language, both tonic and atonic syllables have approximately the same duration, whereas in English, a prototypical stress-timed language, tonic syllables are lengthened and atonic syllables suffer reduction, including (in the case of English) both shorter durations and centralization (Low and Grabe 1995). Rather than being a binary distinction of stress- vs. syllable-timed languages, rhythm is a gradient feature that can

be measured using a series of established rhythm metrics, such as the Pairwise Variability Index (PVI) (Low and Grabe 1995; Grabe and Low 2002). The normalized PVI measurement (nPVI) takes the difference in duration between adjacent segments (in this case, vocalic segments), and divides that value by the mean duration of both vowels, to control for speech rate. Higher nPVI values indicate a more “stress-timed” pattern, whereas lower nPVI values suggest a more “syllable-timed” rhythm. In studies employing nPVI, rhythm has been shown to be susceptible to cross-linguistic influence in bilingual communities, with bilinguals showing intermediate or converged rhythm values (Carter 2005; Shousterman 2014; Carter and Wolford 2016), making prosodic rhythm an important point of inquiry in the development of bilingual lects.

The initial swarm analysis (Ronquest et al. 2020) used the Correlatore 2.3.4 (Mariano 2014) software package to make 59,311 vocalic comparisons across the data set. The results of the linear regression found no significant main effect for any of the independent variables (sex, region of origin, and generation), but some trends in the data point towards possible future change, as G2 speakers showed increased variability as well as a non-significant tendency to produce higher nPVI values (i.e., more “English-like”). Of the seven speakers who produced the highest nPVI values, five were G2 speakers. In this way, Spanish speakers in NC may be showing initial signs of following more established communities in the development of bilingual rhythm.

2.3.4. Bilingual Discourse Markers (DMs)

Discourse markers are “[p]articles that frequently occur in conversation . . . [that] contribute to the overall coherence of discourse by signaling relationships between portions of the speaker’s utterances” (Torres 2011, p. 493). Examples of discourse markers from Spanish and English include *tú sabes*, *o sea*, *entonces*, *como pues*, *you know*, *I mean*, *so*, *like*, and *well*. In situations of language contact, DMs from the contact language are among the first and most common types of lexical borrowings (Torres 2011). For example, Spanish–English bilinguals may use both *so* and *entonces* in the same discourse, either with the same or with differing pragmatic functions. Over time, bilingual speakers may settle on one system of DMs in both languages, as speakers utilize DMs from the dominant language regardless of the language they are speaking. This has led some scholars to take the frequency of bilingual DMs as indicative of the level of integration into the dominant culture (Torres 2002; Lipski 2005). Due to their high frequency and largely subconscious use, bilingual DMs have been described as a “gateway” to other types of borrowings and code switching (Lipski 2005). In established Hispanic communities in the US, the use of English DMs in Spanish is common (50% among New York City Puerto Ricans, Torres 2002; 65% among Chicago Mexi-Ricans, Torres and Potowski 2008; and 68% in New Mexican Spanish, Aaron 2004). One of the striking findings of studies in established Latino communities is the high rates of English DMs even among Spanish-dominant G1 speakers, with English *so* being particularly common across social groups (Lipski 2005; Torres and Potowski 2008).

Following Torres (2002), the initial swarm analysis examined pairs of bilingual DMs that are relatively equivalent across English and Spanish (*you know*~*tú sabes*, *I mean*~*o sea*, *so*~*entonces*, *like*~*como* and *well*~*pues*). Every DM in each interview was coded for the independent social variables age, sex, and generation for a total of 1660 tokens.

The mixed-effects logistic regression (random intercept of speaker) found significant effects of discourse pairs ($p < 0.001$) and generation ($p < 0.001$), with G2 speakers producing significantly more English DMs. Sex approached significance ($p = 0.08$), with women producing more English DMs than men. Overall, the rate of English DMs in NC was much lower than in other areas (11.9% vs. 50% or greater in more established communities), showing an important distinction between the newly forming bilingual community in NC and more traditional Hispanic populations. The low rate of English DMs extended even to *so*, which was found to be ubiquitous among speakers of all generations in other regions of the U.S., with G1 speakers in NC producing less than 1% *so* (vs. *entonces*). Given that *so* has been described as a “core borrowing” (Torres 2002; Torres and Potowski 2008) in

U.S. Spanish, this finding suggests that “NC Spanish may represent an earlier stage of U.S. Spanish development than [more established regions]” (Ronquest et al. 2020, p. 317).

2.3.5. Summary of the Initial Swarm

When each variable was analyzed separately, as in a traditional second-wave study, only DMs differed significantly across generations, with G2 heritage speakers producing significantly more English DMs than G1 immigrants. Nevertheless, the swarm analysis, which was achieved by assessing the individual speaker coefficients associated with random and fixed-effects intercepts (Drager and Hay 2012), revealed important trends that were not apparent in the individual variable analyses. Coefficients (positive or negative) indicated if an individual speaker favored the contact-induced realization for a particular variable (i.e., more occlusive /bdg/, smaller CHA (vowel space), more stress-timed rhythm, and more English DMs). Examination of individual patterns revealed that those who favored the contact-induced realization for three or four of the variables in the swarm tended to be G2 speakers; three of the five speakers who did not favor any contact-induced realizations were G1 speakers. Ronquest et al. (2020) therefore concluded that “in this community, heritage speakers [G2] tend to precede IMs [G1 immigrants] in producing contact-induced realizations” (p. 319).

The patterns evident in the initial swarm also permitted a preliminary assessment of the relationship between variables. The only variable to show a significant main effect of generation (G1 vs. G2) was English DMs, as G1 speakers rarely integrated English DMs in NC, a difference with more established bilingual communities (e.g., Chicago, see Torres and Potowski 2008). Unlike English DMs, both G1 and G2 speakers showed evidence of English-influenced prosodic rhythm, suggesting early adoption of this feature in the development of bilingual lects. Speech rhythm and vowel production are also likely inherently linked, as greater degrees of vowel reduction are apt to result in rhythmic patterns that trend more towards stress timing as well as a smaller vowel space. Results pertaining to lenition of /bdg/ were more complex: Mexican and Central American heritage speakers exhibited opposite patterns, therefore suggesting the potential enhancement of a dialectal feature already present in Central American Spanish (e.g., more occlusive-like productions) and not solely the result of contact with English. In conjunction, Ronquest et al.’s (2020) findings suggest a complex interplay of factors that influence the formation of the linguistic systems of diverse speakers residing within the community.

The present study adds four additional linguistic variables that have been shown to vary in U.S. Spanish to the swarm analysis: subject pronoun expression (SPE), phonological filled pauses (FPs), code switching, and the realization of orthographic and <v>. Growing the swarm with additional variables will offer further insight into how language and dialect contact manifest on different levels within the system, and if, how, when, and why members of the community might integrate these features. The background and motivation for the inclusion of each new variable is detailed below.

2.4. New Swarm Variables

2.4.1. Subject Pronoun Expression (SPE)

Spanish is a pro-drop language, and as such, overt subject pronouns are optional and variably appear based on a variety of morpho-syntactic and discourse-pragmatic factors (Otheguy and Zentella 2012). SPE refers to whether a finite verb appears with (*yo hablo*) or without (*Ø hablo*) an expressed subject pronoun, with both realizations meaning “I speak”. SPE in Spanish varieties, particularly those in contact with other languages, has been described as a “showcase variable” in Hispanic sociolinguistics due to the large number of studies carried out on this phenomenon (Bayley et al. 2012). The abundance of research stems from SPE’s existence at the interface between morpho-syntax and pragmatics, an area particularly susceptible to cross-linguistic influence and/or bilingual effects. Studies have shown that SPE responds to factors such as the person, number, and definiteness of the subject; the tense–mood–aspect (TAM) of the verb, often coded as distinctive (first

person and third person singular verbs have different forms, as in the preterit *hablé* vs. *habló* vs. non-distinctive (first person and third person singular verbs have the same form, as in the imperfect *hablaba*–*hablaba*); the lexical content of the verb (estimative—opinion verbs such as *creer*, stative—verbs not involving any activity such as *ser* and *estar*, external activity—verbs that involve a physical action such as *ir* or *hacer*, or mental activity—verbs of thinking and volition such as *elegir* and *querer*); reflexivity; and switch reference, which refers to whether the subject of the target verb is the same or different than the subject or object of the preceding verb. Switches in reference may be complete (the subjects of the two verbs differ, as in *Mi amiga fue a clase y (yo) fui a la biblioteca* ‘My friend **went** to class and (I) **went** to the library’) or partial (the subject of the target verb is the object of the preceding verb, as in *Mi amiga me dijo que (yo) tenía que estudiar* ‘My friend told **me** that (I) had to study’). Overall, overt subjects are realized more often when the subject or verb is singular, non-distinctive, estimative or referring to a mental activity, non-reflexive, or has a switch in reference from the preceding verb (cf. Sorace 2004; Otheguy and Zentella 2012; Otheguy et al. 2007; among many others for additional information on each of these factors).

Within the United States, many studies have found an increased rate of overt subject pronouns among bilingual populations (Otheguy et al. 2007; Otheguy and Zentella 2012; Abreu 2012; Shin 2013; among others), which may be due to indirect transfer from English (Silva-Corvalán 1994; Otheguy and Zentella 2012; Shin and Otheguy 2009, 2013; Shin 2013; among others). Other studies, however, suggest that observable differences in the surface pronoun rate and in the underlying grammar are due to general processes of bilingual simplification, whereby bilingual speakers lessen their cognitive load by simplifying or weakening underlying grammatical constraints (Sorace 2004, 2005).

In addition to surface effects on pronoun rates, studies have consistently found a weakening of underlying constraints among bilinguals, particularly of G2 and beyond. This weakening of sensitivity to underlying grammatical variables is especially pronounced with regard to switch reference (Silva-Corvalán 1994; Bayley and Pease-Alvarez 1997; Otheguy and Zentella 2012; Shin and Otheguy 2009; Shin 2013). Studies have suggested that a change in sensitivity to switch reference may precede a surface change in overt pronoun rates, as demonstrated by studies in emerging bilingual communities, such as Spanish in metropolitan Atlanta (Limerick 2019), even when there are no significant differences between generations in overall pronoun rate.

2.4.2. Code Switching

Code switching is defined as the “alternating use of two languages within a segment of discourse” (Toribio 2011, p. 532). Despite popular views that code switching is random, chaotic, and reflects a lack of proficiency in one or both of the languages being switched (see examples in Toribio 2004), studies have consistently shown that code switching is structured and rule based (Poplack 1980; Toribio 2004; Anderson and Toribio 2007) and that speakers need a high level of proficiency in both languages in order to produce the most complex, sentence-level code switches (Poplack 1980, 1988). Code switching is common in many, but not all, bilingual communities, and when code switching is accepted within a community, speakers most often code switch as a way of indexing in-group or bi-cultural identities (Myers-Scotton 1995; Zentella 1997). As Zentella (1997, p. 114) concluded, speakers utilize code switches as a “way of saying that they belonged to both worlds, and should not be forced to give up one for the other”.

Code switches can take many forms that vary in their complexity and in the level of bilingualism required (Lipski 2005, 2008; Escobar and Potowski 2015). Individual words from English may be integrated into Spanish discourse by speakers with even the most basic English proficiency. Single-word code switches are distinguished from loanwords based on being spontaneously produced, rather than accepted forms within the community and on the lack of phonological integration for code switches (Escobar and Potowski 2015). Intersentential code switches, which occur between two independent phrases, are the next least complex category of switches, as they do not require speakers to respect the

grammar of both languages (i.e., the equivalence constraint, see [Poplack 1980](#)). In other words, less-proficient bilinguals are able to code switch between sentences “without fear of violating a grammatical rule of either of the languages involved” ([Poplack 1980](#), p. 581). Finally, intrasentential code switches occur at predictable points within the same sentence where the grammar of the two languages is equivalent, generally respecting the integrity of a syntactic constituent ([Poplack 1980](#)). Intrasentential code switches occur only among the most proficient bilinguals ([Poplack 1980](#); [Lipski 2014](#)), and therefore, the presence of more complex code switches can be interpreted as an indicator of the level of bilingualism or linguistic integration of a particular speaker ([Poplack 1980](#)). Examples from our data of each of the three types of code switches considered here are found below:

- Single word switch: *Ella estaba bien surprised* ‘She was very surprised’.
- Intersentential switch: *¿Cómo te puedo decir? The right path* ‘How can I say it (for you)? The right path’.
- Intrasentential switch: *O como cuando van hunting in the woods* ‘Or like when they go hunting in the woods’.

2.4.3. Filled Pauses

Filled pauses (FPs) are “nonsilent hesitations” that serve a variety of functions within conversation, including giving speakers “time to plan utterances and a way to hold the conversational floor” ([Erker and Bruso 2017](#), p. 205). Filled pauses may either be lexical (e.g., English *well, so*,¹ or Spanish *este, sea*) or phonological (e.g., English *u(m)*, or Spanish *e(m)*), where the different default vowels in each language (/ə/ for English, /e/ for Spanish) appear as phonological fillers. The present study focuses on phonological fillers, where there is no overlap with other categories, such as discourse markers (see [Ronquest et al. 2020](#)). FPs have not been widely studied in bilingual U.S. Spanish, with [Erker and Bruso \(2017\)](#) as an important exception. The authors examine three possible realizations of phonological FPs in the Spanish spoken by G1 and G2 in Boston, MA: Spanish [e], English [ə], and a third form, [a]. They find that, as contact with English increases, so do instances of both [ə] and [a]. While [ə] is interpreted as transfer from or convergence with English, [a] is seen as a possible intermediate form that is shared by the phonological systems of both languages, and therefore, may be used by bilingual speakers as a compromise form, lessening the cognitive load of maintaining two separate FP systems ([Erker and Bruso 2017](#), p. 238). Alternatively, the authors note that in usage-based theories, the increase in [a] may stem from the increased presence of [ə] variants in bilingual speech, and speakers “might establish connections between exemplars of this category and those of the vowel category that is most like it acoustically and that is used to similar effect when speaking in Spanish, namely, [a] (and not [e])” ([Erker and Bruso 2017](#), p. 239). Whatever the theoretical model, a decrease in the use of [e] FPs is taken to indicate higher levels of integration into the English-speaking environment.

2.4.4. The Realization of Orthographic b~v

It is widely reported that monolingual varieties of Spanish do not distinguish between orthographic and <v>, as both are pronounced as variants of the phoneme /b/ ([Hualde 2005](#)). Studies of bilingual Spanish in the United States, however, have reported a distinction between bilabial /b/ corresponding to and labiodental /v/ corresponding to <v>, particularly among G2 speakers ([Trovato 2017](#)). Given that English is characterized by a phonemic distinction between /b/ and /v/, the existence of an incipient distinction among U.S. Spanish speakers is taken to be due to the influence of orthographic <v> in English ([Rao and Ronquest 2015](#); [Boomershine and Ronquest 2019](#)). Among speakers with less exposure to Spanish, [Rao \(2014\)](#) found more “tense” (i.e., less monolingual-like) pronunciations of orthographic <v> compared to orthographic , although the lack of fricatives in his data made him doubt the direct influence of English fricative /v/.² In a more detailed analysis, [Rao \(2015\)](#) does allow for the influence of English on orthographic <v>, stating that “[t]he relatively high rates of [tense approximant] realizations associated

with increases in articulatory tension that trend toward a fricative production suggest that the English /v/ phoneme interfered with productions of /b/, which was exacerbated by seeing <v>” (p. 68).

While a few studies have undertaken acoustic analyses of Spanish and <v>, researchers are not in clear agreement on which acoustic measures best capture the difference between bilabial and labiodental realizations. Trovato’s (2017) study of the production and perception of and <v> among bilinguals residing in Texas revealed a significant effect of segment duration and intensity difference, but not spectral center of gravity (COG), in spite of COG being one of the primary acoustic correlates utilized to distinguish fricatives. COG is a weighted mean of frequencies measured in Hertz (Hz) that indicates how high or low in the spectrum most of the energy is concentrated (Boersma and Weenink 2022). Different from Trovato (2017), Chetty (2018), who incorporated videos of lip movements, found that COG was the best predictor of labiodentalization: the presence of teeth in the video correlated with significantly higher COG values, suggesting higher rates of frication, as fricatives are produced with the bulk of their energy at higher frequencies (Ladefoged and Disner 2012). Likewise, initial analyses of the present data confirm that COG is a more reliable predictor of grapheme, as the COG of productions of <v> trends higher than that of , while duration and intensity difference did not yield significant results. Based on these results, our analysis focuses on COG, although further exploration of the relationship between /<v> and various acoustic properties is warranted in future studies.

2.4.5. Growing the Swarm

The addition of the four variables outlined above, in conjunction with the four variables included in the initial swarm analysis (Ronquest et al. 2020) contributes to our understanding of how bilingual communities and individual speakers integrate contact-induced forms in the early stages of contact, and the results presented here will focus on differences between G1 and G2 speakers.

To facilitate comprehension of the new swarm variables and how they connect with those analyzed in the initial swarm (Ronquest et al. 2020), we begin with a review of the general methodology, followed by methods specific to the analysis of each individual variable (SPE, code switching, FPs and <b/v>). Separate results sections for each variable are presented next, and the new integrated swarm analysis including all eight variables is presented in Section 5.

3. Materials and Methods

3.1. General Methods

The data for the present study were obtained from sociolinguistic interviews with 10 men and 13 women (23 informants total) of Hispanic/Latino descent ranging in age from 20 to 53 (average age 27.3 years). Informants were further subdivided into two groups: G1 speakers who were foreign-born immigrants (n = 12) and G2 speakers (n = 11) who were born in the United States or born outside of the United States and immigrated by the age of three. Sixteen participants were of Mexican Heritage and seven were from Central America (El Salvador and Guatemala). Detailed information regarding demographics can be viewed in Appendix A.

Informants participated in sociolinguistic interviews that lasted between 30 and 60 min and included questions pertaining to general experiences living in the United States and abroad, customs, and similarities and differences between the U.S. and their heritage countries. Interviews were recorded in quiet locations utilizing Zoom H2 digital recorders (44.1 kHz, 16 bit) and subsequently transcribed orthographically in Praat (Boersma and Weenink 2022) to facilitate forced alignment with the *Forced Alignment System for Español* (FASE; Wilbanks 2015). Alignment facilitated acoustic analysis of phonetic variables in the first swarm (i.e., lenition of /bdg/, vowel space, nPVI) and acoustic properties of and <v> in the current study. The automated system significantly increased the number of

analyzable units by orders of magnitude (for example, more than 125,000 vowel tokens in the initial swarm and more than 3500 <b/v> tokens here; see [Labov et al. 2013](#)).³

Additional details regarding the specific coding scheme and statistical analysis for each of the four variables included in the present analysis are provided below. The swarm analysis, which focuses on analyzing individual speaker patterns and the interactions between all variables, is presented in Section 5.

3.2. Methodology: Subject Pronoun Realization (SPE)

For each of the sociolinguistic interviews, the first 100 finite verbs that fit into the envelope of variation were identified. The requirements for inclusion in the envelope of variation were based on the coding manual found in [Otheguy and Zentella \(2012\)](#) and included all finite verbs that could appear with an overt subject pronoun, whether or not an overt subject pronoun was present. Verbs appearing with lexical subjects, inanimate subjects, and non-personal pronouns (such as *eso* or *aquel*) were considered outside the envelope of variation and thus excluded from the study.

A total of 2265 tokens were included in the analysis. A series of mixed-effects logistic regression models was run in Rbrul ([Johnson 2009](#)), with a random intercept of speaker. The binary-dependent variable was whether the pronominal subject was null or overt, and independent variables included a variety of morpho-syntactic, pragmatic, and social factors found to be important predictors of SPE in previous studies ([Otheguy and Zentella 2012](#); [Otheguy et al. 2007](#); among many others), as well as speaker generation, sex, and region of origin, as seen in Table 1.

Table 1. Variables included in the analysis of SPE.

Variable	Levels
Dependent	
SPE	Overt Null
Independent	
	Yo
	Tú (definite)
	Tú (indefinite)
Person/number/definiteness	Él/ella (definite)
	Nosotros/nosotras
	Ellos/ellas (definite)
	Ellos/ellas (indefinite)
Distinctiveness of TAM	Distinctive
	Non-distinctive
Switch reference	Complete switch
	Partial switch
	No switch
	External activity
Lexical content	Mental activity
	Estimative
	Stative
Generation	G1
	G2
Sex	Male
	Female
Region of origin	Mexico
	Central America
Speaker	Random intercept

In order to establish the comparative constraint hierarchies across groups ([Tagliamonte 2011](#)), the relative importance of each linguistic variable was determined based on model comparison via AIC, with one independent variable removed for each of the subsequent

models.⁴ The AIC values for each model were then compared, with the size of the difference in AIC value between the full model and the reduced model indicating the strength of each variable in the analysis (see [Kapatsinski 2012](#)).

3.3. Methodology: Code Switching (CS)

In each sociolinguistic interview, all code switches were identified and categorized as single word, intersentential or intrasentential, for a total of 779 instances of code switching. Observed patterns in the data were confirmed via a chi-square analysis run in R ([R Core Team 2022](#)), as well as mixed-effects logistic regressions in Rbrul ([Johnson 2009](#)) with type of switch as the dependent variable and a random intercept of speaker. Two separate regression models were run, one with sentence-level switches (vs. word-level switches) as the application value, and a second with the most complex type of switch (intrasentential vs. other) as the application value. The independent variables for both models were speaker generation, sex, and region of origin. Table 2 presents the variables included in the analysis.

Table 2. Variables included in the analysis of code switching.

Variable	Levels
Dependent	
Code-switch type	Sentence-level code switching (Intrasentential + intersentential) (analysis 1) Word Intrasentential (analysis 2) Intersentential + Word
Independent	
Generation	G1 G2
Sex	Male Female
Region of origin	Mexico Central America
Speaker	Random intercept

3.4. Methodology: Filled Pauses (FP)

For this initial analysis, every phonological filled pause in each interview was impressionistically coded as a variant of [e(m)], [a(m)] or [ə(m)] based on the vowel heard in each token.⁵ A total of 1925 filled pauses were included in the analysis, and a chi-square analysis was conducted in R ([R Core Team 2022](#)) to determine the significance of the overall distribution of forms across generations. For the mixed-effects logistic regression analysis fit to the data in Rbrul ([Johnson 2009](#)), the binary-dependent variable was the bilingual forms [ə] + [a] vs. the monolingual Spanish form [e]. The independent variables were speaker generation, sex, and region of origin. Speaker was included as a random intercept. The variables included in the analysis of filled pauses are presented in Table 3.

Table 3. Variables included in the analysis of FPs.

Variable	Levels
Dependent	
Filled pause vowel	[ə] + [a] [e]
Independent	
Generation	G1 G2
Sex	Male Female
Region of origin	Mexico Central America
Speaker	Random Intercept

3.5. Methodology: The Realization of Orthographic <b/v>

In each of the sociolinguistic interviews, all instances of intervocalic /b/ were identified, for a total of 3550 tokens. Following [File-Muriel and Brown \(2011\)](#), a Praat ([Boersma and Weenink 2022](#)) script was used to measure the COG of the middle 60% of the token, with the outer 40% being ignored in order to minimize any effect of surrounding phonemes on the COG measurement. Since initial analyses determined that the data were highly skewed to the right,⁶ the COG measurements were log transformed in order to normalize the distribution of the data for analysis. Two mixed-effects linear regression models were fit to the data in Rbrul ([Johnson 2009](#)) with random intercepts of speaker and word. The dependent variable of the first was log (COG), and the independent variables were the following vowel and word position (initial vs. medial), grapheme (vs. <v>), as well as speaker generation, sex, and region of origin (see [Table 4](#)). Finally, in order to better compare <b/v> patterns across speakers, the mean COG value for was subtracted from the mean COG value for <v> for each speaker, with a positive value indicating a higher mean COG for orthographic <v>. The difference in mean COG values formed the dependent variable of the second linear regression model, with speaker as a fixed effect. The results of the COG difference analysis will be discussed as part of the swarm analysis in [Section 5.1](#).

Table 4. Variables included in the analysis of <b/v>.

Variable	Levels
Dependent	
COG	Continuous variable (analysis 1) COG difference <v>— (analysis 2)
Independent	
Following vowel	[i] [e] [a] [o] [u]
Word position	Initial Medial
Grapheme	 <v>
Generation	G1 G2
Sex	Male Female
Region of origin	Mexico Central America
Speaker	Random intercept
Word	Random intercept

4. Results

4.1. Subject Pronoun Expression (SPE)

Table 5 presents the results of the mixed-effects logistic regression. In the regression table, a positive log odds indicates a favoring of an overt subject pronoun. The overall analysis with all speakers found significant main effects of person, number and definiteness, switch reference, lexical content of the verb, and reflexivity, with patterns largely matching those found in previous studies (Table 5). Generation approached significance; however, a comparison of overall pronoun rates shows very little difference between G1 and G2 speakers in the present data. This corroborates the findings of Limerick (2019) for Roswell, Georgia, another community in the Southeast, further supporting that the southeastern U.S. may be at an earlier stage of bilingual SPE development than more established communities, where significant differences across generations are frequently found (e.g., New York City, see Otheguy and Zentella 2012).

Table 5. Results of the multivariate one-level mixed-effects regression model, speaker as a random factor. * Significant factors with a p -value < 0.05.

Factor	Log Odds	N	% Overt Pro	p-Value			
<i>Person/Number and Definiteness</i>				<0.001 *			
él-ella.definite	1.902	267	38.2				
1st singular—yo	0.847	1214	21.6				
3rd plural—ellos-ellas.definite	0.567	209	18.2				
2nd singular—tú.definite	−0.352	58	8.6				
2nd singular—tú.indefinite	−0.494	151	6.6				
1st plural—nosotros-nosotras	−0.589	292	8.6				
3rd plural—ellos-ellas.indefinite	−1.881	74	2.7				
<i>Switch Reference</i>				<0.001 *			
Complete switch	0.643	945	25.5				
Partial switch	−0.284	181	18.2				
No switch	−0.359	1139	15.0				
<i>Reflexivity</i>				<0.001 *			
Non-reflexive	0.392	1997	20.7				
Reflexive	−0.392	268	11.6				
<i>Distinctiveness of TAM</i>				<0.001 *			
Non-distinctive	0.248	475	26.1				
Distinctive	−0.248	1790	17.9				
<i>Lexical Content</i>				0.002 *			
Estimative	0.676	99	43.4				
Stative	−0.132	562	22.0				
External activity	−0.220	1229	17.4				
Mental activity	−0.324	375	17.1				
<i>Generation</i>				0.094			
G1	0.238	1173	19.9				
G2	−0.238	1092	19.2				
<i>Sex</i>				0.24			
Female	0.173	1282	21.6				
Male	−0.173	983	17				
<i>Region</i>				0.60			
Central America	0.08	686	24.2				
Mexico	−0.08	1579	17.6				
N	df	intercept	overall prop	AIC	R2.fixed	R2.random	R2.total
2265	18	−2.348	0.196	1970.175	0.224	0.063	0.287

Table 6 shows the comparative constraint hierarchies for speaker subgroups when analyzed separately, broken down by origin and generation, revealing important differences

between subgroups (Full regression tables are found in Appendix B). G1 speakers show the same constraint hierarchies, indicating that they share an underlying grammar for SPE, regardless of their region of origin. All five linguistic variables were significant for Mexican G1 speakers, whereas only the first three variables were significant for Central American G1 speakers.⁷

Table 6. Variable hierarchies by generation and origin. * Significant factors with a p -value <0.05 .

Mexican G1	Central American G1	Mexican G2	Central American G2
1. Person/Number *	1. Person/Number *	1. Person/Number *	1. Person/Number *
2. Switch *	2. Switch *	2. Lexical Content *	2. Switch *
3. Lexical Content *	3. Lexical Content *	3. TMA Distinctive *	3. Lexical Content
4. TMA Distinctive *	4. TMA Distinctive	4. Reflexivity *	4. TMA Distinctive
5. Reflexivity *	5. Reflexivity	5. Switch *	5. Reflexivity

Several important differences appear when comparing G1 and G2 speakers. Mexican G2 speakers show the same overall hierarchy and significance as Mexican G1 speakers, except that switch reference is much less important for the G2 group. This finding suggests that these speakers show a decreased sensitivity to switch reference, which has also been reported in contact varieties of Spanish around the world (Shin and Otheguy 2009; Michnowicz 2015). Central American G2 speakers show the same order of constraints as their G1 counterparts, but with fewer significant differences (i.e., lexical content of the verb is no longer significant for G2 speakers).

In summary, while there are very few differences in the use of overt subject pronouns across generations overall, a comparative analysis of constraint hierarchies reveals a simplification of the underlying grammar for G2 speakers, realized as a weakening of switch reference or a reduction in the number of significant predictors, which suggests possible future changes in rates of overtly expressed subject pronouns in NC Spanish.

4.2. Code Switching

Of the 779 total code switches in the corpus, the vast majority were single-word switches, as has also been observed in previous studies (Poplack 1980). Intersentential switches were a distant second (7%), with intrasentential switches the least frequent and comprising only 2.6% of the tokens (Table 7). In this way, the frequency of code-switch types in the corpus reflects the complexity of each type of switch, as simpler switch types are also more frequent in the data.

Table 7. Frequency of code-switch types.

Switch Type	N	% of Data
Single word	702	90.1%
Intersentential	57	7.3%
Intrasentential	20	2.6%

The distribution of code-switch types across generations follows the expected pattern, with 19/20 (95%) of the most complex intrasentential code switches appearing among G2 speakers. The chi-square analysis revealed a significantly different pattern across generations ($\chi^2(2, N = 779) = 8.1604, p = 0.017$), as seen in Figure 1, which presents code-switch (CS) type by generation. The width of the box for each generation indicates the number of tokens, showing that G2 speakers produced many more code switches overall, and likewise produced the majority of inter- and intrasentential code switches.

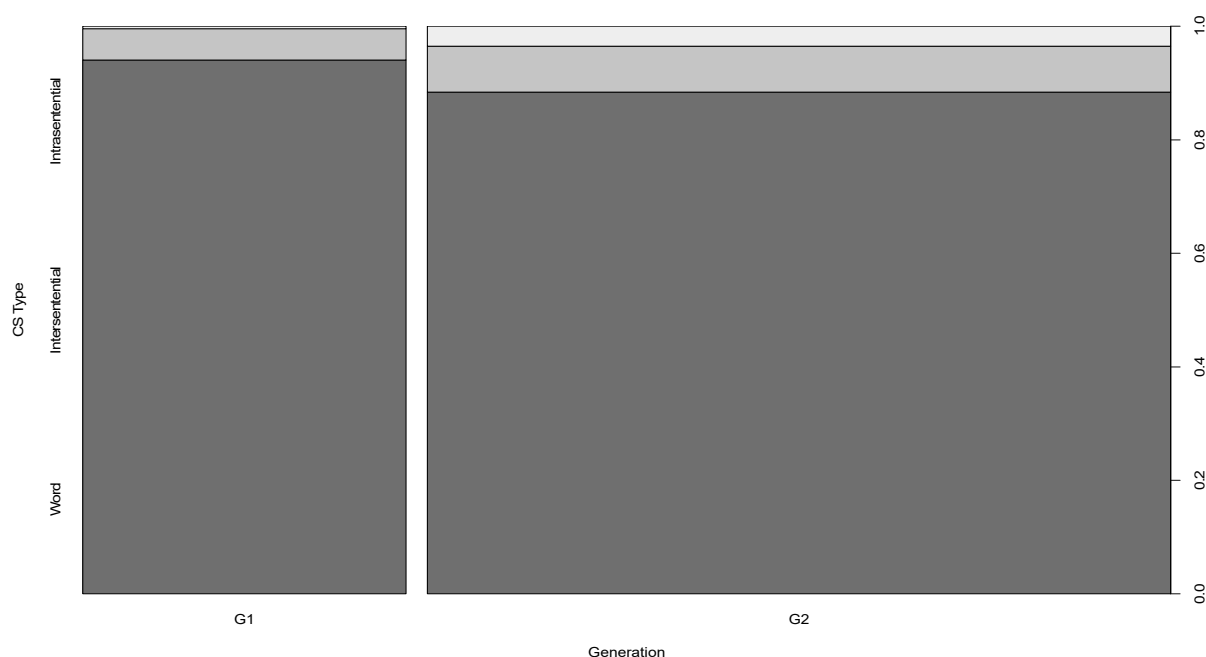


Figure 1. Code switch type by generation.

The mixed-effects logistic regression comparing sentential switches vs. single word revealed a significant effect of generation ($p = 0.046$), with G2 speakers producing significantly more sentential switches than G1. When comparing the most complex switch type (intrasentential switches) vs. other switch types (intersentential + word-level), the difference between generations was even more pronounced ($p = 0.025$), as only one of the twenty intrasentential code switches was produced by G1 speakers. Region of origin and speaker sex were not significant predictors of switch type. In summary, G2 speakers in NC produced more code switches overall than G1 speakers, a difference that was most pronounced for sentential-level code switches, suggesting the potential for increased levels of code switching among future generations of speakers.

4.3. Filled Pauses (FPs)

The analysis of FPs revealed that G2 speakers produced more than twice as many phonological filled pauses as G1 speakers (1304 vs. 621), which may suggest differences in fluency (see [García-Amaya 2009](#) and sources therein).⁸ Regarding specific forms (see Figure 2), G2 participants produced almost three times as many instances of [ə] (28% compared to 11% for G1), suggesting increased English influence in their Spanish. Conversely, G1 speakers showed higher rates of [e] (20% vs. 15% for G2). Speakers of both generations showed [a] as the majority form (69% G1, 57% G2), providing further evidence in support of [Erker and Bruso's \(2017\)](#) claim that [a] may serve as an intermediate form between Spanish [e] and English [ə]. A chi-square analysis found the overall distribution of filled pause variants across generations to be significant ($X^2(2, N = 1925) = 74.612, p < 0.001$).

The mixed-effects logistic regression comparing the bilingual forms [ə] and [a] to the monolingual variant [e] found a marginally significant effect of sex ($p = 0.053$), with men producing more bilingual FPs than women (93% vs. 79%). Generation approached significance ($p = 0.09$), with G2 speakers producing more bilingual FPs than G1 speakers (85% vs. 80%). Region was not a significant predictor ($p = 0.21$).

Despite the observed differences across generations, further analysis of the R^2 values for fixed and random factors indicates that individual speaker differences are the primary factor in FP variation in NC, as the random factor of speaker accounts for 43% of the observed variation compared to only 12% for the fixed effects. This is clearly seen in Figure 2, which shows the filled pause variants for each speaker divided by generation.

With the exception of one speaker (2014-13, a G1 Salvadoran male who has spent 19 years in the U.S.), G2 speakers show much greater variation than G1, both in the number of FPs and in their realization. Likewise, while most G2 speakers produced at least some instances of /ə/, the bulk of the centralized tokens was produced by two speakers (2012-15, 2013-19, a female of Salvadoran heritage and a male of Mexican heritage, respectively; both speakers report speaking English and Spanish at home). At the same time, Figure 2 shows that most speakers of both generations display intra-speaker variation, even if a majority has settled on /a/ as their default FP. These findings will be addressed further in the discussion.

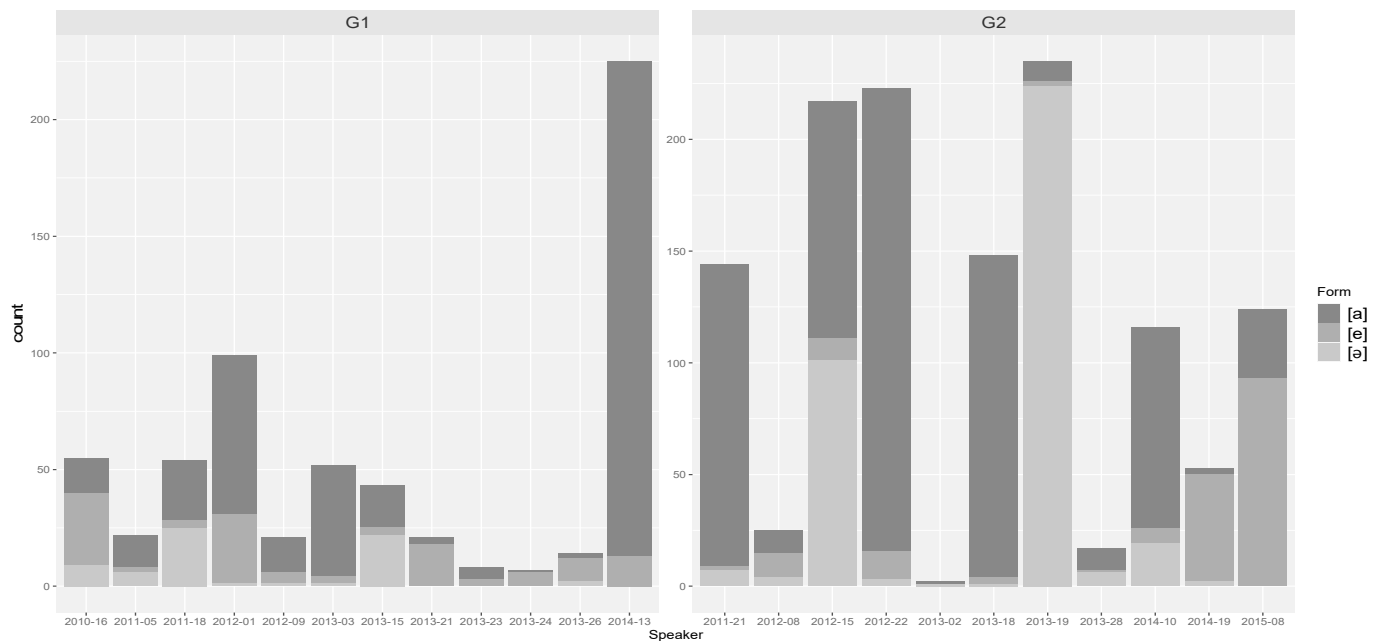


Figure 2. Filled pause variants by speaker and generation.

4.4. The Realization of Orthographic <b/v>

The mixed-effects linear regression model revealed a significant main effect of following vowel ($p < 0.001$), with grapheme being marginally significant ($p = 0.051$). COG values were higher for the grapheme <v>, suggesting a more fricative, labiodental pronunciation (Chetty 2018). Word position, speaker sex, generation, and region of origin were not significant predictors of COG in the present analysis.⁹ A visual inspection of the data shows that both generations of speakers essentially show the same pattern of higher COG values for orthographic <v> than for , which previous research has shown may indicate a fricative, labiodental articulation (Chetty 2018). The results by generation and grapheme are seen in Figure 3.

While the overall pattern is similar, the boxplots in Figure 3 suggest an increased separation between and <v> for G2 speakers, as evidenced by less overlap in the boxes between and <v>, as well as greater differences in the median values across graphemes. This observation is borne out by the results of an additional mixed-effects linear regression model that included an interaction of generation and grapheme, which indicated a slight tendency for G2 speakers to make a greater distinction between orthographic and <v> in NC ($p = 0.085$).

In sum, even with a relatively small number of speakers included in the analysis, the potential for both individual variation and the impact of surrounding phonetic context, our analyses indicate a possible generational difference with regard to the articulatory and acoustic properties of and <v> among NC bilinguals.

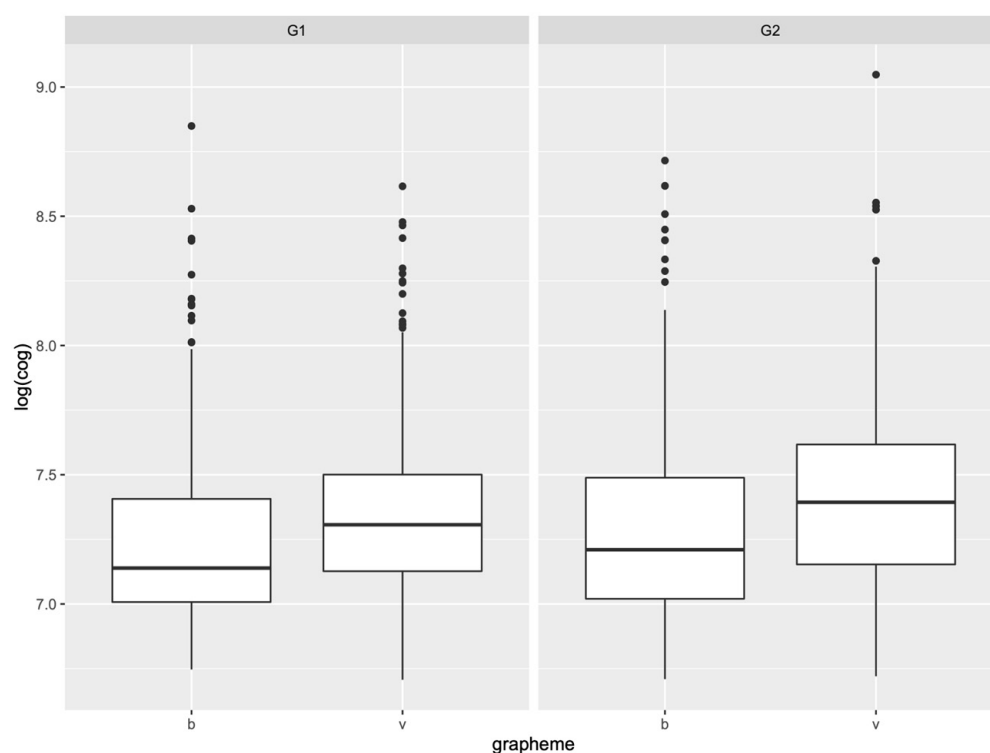


Figure 3. Log (COG) by grapheme and generation.

5. Discussion: The Swarm Analysis

5.1. Speaker Rankings

One of the primary benefits of a swarm analysis is the ability to examine how variables pattern both across and within individual speakers. Here, the results of the four linguistic variables from the present study (SPE, code switching, FPs, and <b/v>) are added to the four initial swarm variables (/bdg/ lention, vowel space via CHA, prosodic rhythm via nPVI, and bilingual DMs) analyzed in [Ronquest et al. \(2020\)](#) across the same group of speakers. A focus on individual speaker variation can provide additional insight into the creation of (bilingual) linguistic norms in a “New Destination” community such as NC. As in [Ronquest et al. \(2020\)](#), we follow [Drager and Hay \(2012\)](#), who show how the random effects coefficients from mixed-effects models can be used to make comparisons between individual speakers by using the individual speaker estimates from each model to rank speakers as favoring (positive coefficient) or disfavoring (negative coefficient) a particular realization for each variable. The random effects intercepts were utilized for FPs, DMs, SPE, /bdg/, and code switching. Since the analyses of nPVI, CHA, and <b/v> COG difference produced only one mean value per speaker, the fixed-effects estimates for speaker (run in separate models) were used to create the ranking for these variables. A positive coefficient indicated that a speaker favored the variant that could be considered “contact-induced” or “less monolingual”: higher nPVI, FPs with a vowel other than [e], more English DMs, more overt subject pronouns, a more occlusive realization of /bdg/, higher rates of intrasentential code switching, and higher COG values for <v>, suggesting more labiodental variants. For CHA, the expected bilingual pattern is a smaller vowel space ([Menke and Face 2010](#)), which was indicated by a negative coefficient.

In Figure 4, a check mark indicates that a speaker statistically favored the contact-induced variant for that variable. Speakers were then ranked according to the number of contact variants they favored. For example, speaker 2011-21 (first row) statistically favored higher nPVI, bilingual FPs, English DMs, overt subject pronouns, greater intensity differences for /bdg/, a smaller vowel space and intrasentential code switching. This speaker did not favor higher COG values for <v> than for .

Speaker	Generation	Region	Sex	nPVI	Filled Pauses not /e/	DMs	SPE	/bdg/	CHA	IntraCS	b/v.Diff
2011-21	G2	CAm	Female	✓	✓	✓	✓	✓	✓	✓	
2012-22	G2	CAm	Female	✓	✓	✓	✓		✓	✓	✓
2013-19	G2	Mex	Male	✓	✓	✓	✓	✓	✓		
2011-05	G1	Mex	Female	✓	✓	✓	✓		✓		
2012-09	G1	Mex	Female	✓	✓	✓	✓				✓
2012-15	G2	CAm	Female	✓	✓			✓	✓		✓
2013-15	G1	Mex	Male		✓	✓		✓		✓	✓
2013-18	G2	Mex	Female	✓	✓	✓			✓		✓
2012-01	G1	Mex	Female	✓	✓	✓		✓			
2013-03	G1	Mex	Male	✓	✓		✓	✓			
2013-21	G1	Mex	Female	✓				✓	✓		✓
2014-10	G2	CAm	Female	✓	✓	✓			✓		
2014-19	G2	Mex	Female	✓			✓			✓	✓
2011-18	G1	Mex	Male		✓		✓				✓
2013-02	G2	Mex	Male		✓						✓
2013-23	G1	Mex	Male				✓	✓			✓
2014-13	G1	CAm	Male	✓	✓						✓
2015-08	G2	Mex	Female	✓		✓			✓		
2013-28	G2	Mex	Male			✓				✓	
2012-08	G2	CAm	Female					✓			
2013-24	G1	Mex	Male	✓							
2010-16	G1	CAm	Female								
2013-26	G1	Mex	Male								

Note. A check mark indicate the speaker favored the contact-induced realization for that specific variable.

N of speakers favoring the	G1	7	7	4	5	5	2	1	6
contact form by generation	G2	8	7	7	4	4	7	4	5

Figure 4. The variable swarm variables by speaker. A check mark indicates that the speaker statistically favors the contact-induced variant.

Although the speaker rankings in Figure 4 indicate that the preference for bilingual forms in NC Spanish is highly variable, some important trends emerge when analyzing individual speakers. First, while no speaker statistically favors contact forms for all eight variables, the two speakers who favor contact forms for seven out of eight variables are both G2 (as indicated by checkmarks in Figure 4). Likewise, of the eight speakers who favor contact forms for five or more of the variables, five are G2. At the bottom of the chart, of the four speakers who favor one or fewer contact forms, three are G1. These results are similar to those for the initial swarm analysis in Ronquest et al. (2020) and speak to the robustness of these patterns even as more variables are added. The middle of the chart is characterized by substantial variation with respect to generation, as predicted by theories of new dialect formation and koineization (Kerswill 2013), as Spanish speakers in NC negotiate the newly forming norms in their community.

A comparison of patterns of use across G1 and G2 speakers reveals further detail regarding how bilingual forms are integrated into the community. The numbers at the bottom of Figure 4 indicate how many speakers of each generation favor the contact-induced form for each variable. A majority of the variables show a relative balance between G1 and G2 speakers, while three variables show a clear effect of generation: English DMs (seven G2 speakers favoring vs. four G1 speakers), smaller CHA (seven G2 vs. two G1), and intrasentential code switching (four G2 vs. one G1). These patterns can suggest an order of integration of bilingual forms, with English-like rhythm, FPs, SPE, occlusive [bdg] and labiodental productions of <v> appearing in the speech of G1 speakers, whereas English DMs, more complex code switches, and changes to the vowel space appear to require a higher level of English dominance in order to enter into the speech of NC Spanish speakers.

One interesting observation is that the bilingual forms that involve the direct integration of English words and phrases (DMs, code switching), and therefore perhaps are most salient to speakers, are also among the last to be integrated. While an interviewer effect has almost certainly played a role in suppressing code switching among some participants, as most interviews were conducted by L2 speakers from outside the Latino community, speakers' attitudes toward bilingual speech may also be an important factor. Studies have repeatedly demonstrated that many speakers hold negative attitudes toward overtly bilin-

gual forms, such as code switches (Anderson and Toribio 2007; Rangel et al. 2015; Mata 2022), and anecdotal evidence from our own outreach efforts with the Latino community in NC finds similar strong, negative reactions toward what some speakers interpret to be “Spanglish.” These commonly held negative attitudes toward “mixed” speech may be reflected in speakers’ reluctance to integrate overtly English forms into their speech, at least in the context of a sociolinguistic interview. Erker (2017) argues that highly salient variables are sites of convergence and overt manipulation, as speakers’ conscious awareness of these features is subjected to social forces within the community. On the other hand, less-salient variables, such as many phonetic or morpho-syntactic features (e.g., SPE), respond primarily to the pressures of cognitive economy “that comes at little social cost” (Erker 2017, p. 16). The present findings reinforce this possibility, and give further weight to arguments based on salience as a deciding factor in the adoption of a feature in a bilingual community.

In light of this possibility, one interesting question that arises is why do some speakers choose to utilize overtly contact-induced forms in an interview context while others do not. For some speakers, particularly of G1, the lack of code switching and other bilingual forms in their interviews may very well reflect their normal linguistic patterns, as it has been argued that “New Destination” communities such as NC lack the critical mass of bilingual speakers required to encourage the wide-spread adoption of many contact variants (Ronquest et al. 2020). On the other hand, variationist studies in the “third wave” tradition (Eckert 2005, 2012; Mendoza-Denton 2002, 2010) emphasize the social agency of speakers to “actively [construct their identities] as they creatively and aesthetically combine linguistic elements” (Mendoza-Denton 2010, p. 189). In other words, the question becomes why have some speakers opted to utilize overtly bilingual forms (such as code switching) during a semi-formal sociolinguistic interview with an outsider? Woolard (1999) argues that bilingual speakers have the option of bivalency, “the use by a bilingual of words or segments that could ‘belong’ equally, descriptively and even prescriptively, to both codes” (p. 7). Michnowicz et al. (2018) have made similar arguments for English loanwords in NC Spanish, as speakers who use semantic calques such as “carpeta” (*carpet*, “rug” based on English “carpet” rather than Spanish “alfombra”) have made a (semi-) conscious choice to remain in Spanish mode, instead of switching to English “carpet”. In the same way, speakers who choose to integrate code switches or English DMs into their Spanish may be actively choosing to use their (bilingual) variety of Spanish, rather than simply switching to English, a language that many G2 speakers may be more comfortable using. One hypothesis would be that speakers who favor contact forms for these salient variables are actively indexing a bilingual Latino identity, showing themselves to be part of the newly forming bilingual community in NC. Although the present data do not permit us to make definitive conclusions with regard to the role of speaker agency in how bilinguals choose to employ (or not) bilingual codes, the (arguably) more salient variables such as DMs and code switching can be most overtly and easily manipulated by speakers in an interview context and could therefore be viewed as tools for identity construction.

5.2. Correlations across Variables

In addition to speaker-specific patterns of use and integration, the swarm analysis also allows for an examination of the correlations between variables. Following research in dialectology that examines the co-occurrence of variables across regional dialects (Coloma 2012), correlations between variables were run on each generation separately, and were plotted using the *corrplot* package (Wei and Simko 2021) in R (R Core Team 2022). Significance of correlations was determined with *rquery.cormat*.¹⁰

Close inspection of Figure 5 reveals several notable differences between variable correlations across generations. First, assessment of the statistical relationships among variables within G1 speakers indicates that there is only one moderately strong correlation: a positive correlation between FPs and DMs. The significant correlation between FPs and DMs ($p = 0.04$) indicates that as G1 speakers use more bilingual FPs ([ə] or [a]), they also use more English DMs (*you know, I mean, so, like, and well*). Given that FPs and DMs

serve similar discourse functions, such as turn taking and holding the floor (see [Erker and Bruso 2017](#)), the correlation between the two variables is not surprising and supports the finding that if G1 speakers integrate FPs and DMs into their Spanish, they tend to integrate them together.

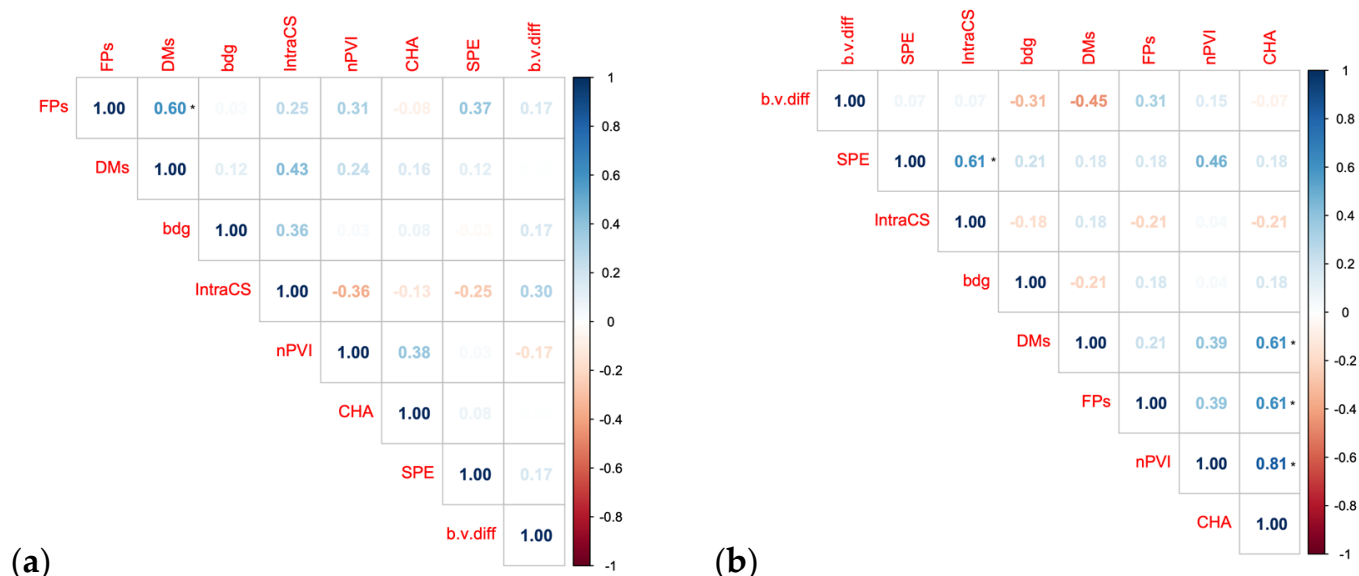


Figure 5. Correlations between variables by generation. * indicates a significant correlation at the $p \leq 0.05$ level. (a) G1 speakers; (b) G2 speakers.

G2 speakers show a higher rate of correlation, with four significant moderate or strong correlations compared to only one among G1 speakers. Several of these correlations are structurally connected; for example, as vowel space (CHA) shrinks due to increased centralization, rhythm values as shown by nPVI will also show a more stress-timed pattern, as predicted in [Ronquest et al. \(2020\)](#). Likewise, CHA also correlates with centralized FPs, again indicating the systematicity of bilingualism, as a change in one variable has impacts throughout the linguistic system(s). Interestingly, the number of speakers favoring bilingual forms for each of these variables (15 for nPVI, 14 for FPs, 9 for CHA) suggests that changes in prosodic rhythm and FPs act as a gateway and precede overall changes in vowel space, although we would hypothesize that once these processes begin, they likely feed off of one another.

The correlation between SPE and code switching shows how these two variables are integrated in tandem, as they both lie at the morpho-syntactic/pragmatic interface that has been shown to be particularly susceptible to cross-linguistic influence ([Sorace 2004](#)). Other connections are less obvious, such as the significant correlation between English DMs and CHA, while the correlation between DMs and FPs, significant among G1 speakers, has lost strength among G2 participants. An examination of bilingual DM and FP rates across speakers and generation provides an explanation. As seen in [Figure 6](#), among G1 speakers, English DMs only appear among the speakers with the highest rates of bilingual FPs, thereby producing a significant correlation. Among G2 speakers, however, English DMs are present for all but one speaker, regardless of their level of bilingual FPs, and the rate of English DMs does not rise in tandem with rates of bilingual FPs, thereby weakening the correlation. As observed in [Section 4.3](#), inter-speaker variation is an important factor in FPs, and both generations exhibit vastly different rates across speakers, although to a greater extent among G1.

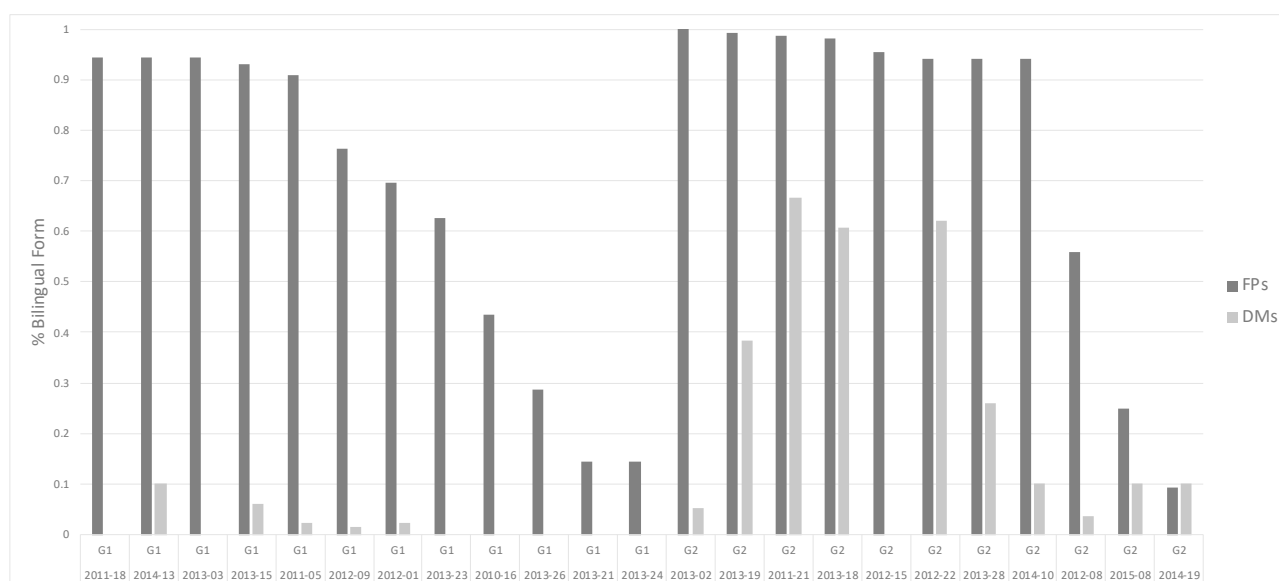


Figure 6. Rates of contact-induced FPs and DMs by generation and speaker.

In sum, connections showing how contact variables are not integrated in isolation only become apparent through a swarm approach. Furthermore, the swarm analysis and correlations between variables offer further insight into when, how, and why contact-induced forms are integrated into the bilingual system. The swarm approach combines a series of traditional, second-wave style studies on individual variables and applies a third-wave style focus on the behavior of individual speakers and variables as part of a larger, emerging bilingual system. In this way, a swarm analysis not only is able to provide a detailed panoramic picture of how variables pattern and interact within and across speakers, but is also crucial for identifying order of integration of features, which can serve as a springboard for future, more “traditional” third-wave studies.

6. Conclusions

By expanding previous research on a variable swarm in Spanish in NC (Ronquest et al. 2020), the present study allows for a more fine-grained approach to understanding language variation and the formation of bilingual norms in a newly forming community. The most important insights from the present study, including patterns of variable use across speakers and the correlations between variables that allow for a proposed order of integration of bilingual forms, would not have been possible without a swarm analysis. Additionally, the observation that English DMs and code switching—the most overt and salient strategies of English integration—are adopted later than phonetic traits, such as prosodic rhythm and vowel space modifications, speaks to the agency of individual speakers to use or not use particular variables not only as a means to index their identities, but also as a part of the process of active identity construction (Mendoza-Denton 2010). It is precisely the salience of these variables that makes them available for conscious manipulation (Erker 2017), and speakers who actively choose to employ stigmatized variants (such as code switches) in their Spanish may do so to set themselves apart from less integrated speakers in the community (see Zentella 1997). In this way, what is perceived as heightened “English influence” may actually be the actively created reflection of a bilingual/bicultural identity. The connection between contact forms and the indexation of identity was explicitly noted by Michnowicz et al. (2018) for English loanwords in NC Spanish, and the inter- and intra-speaker variation observed in the present data suggests that a complex interplay of social networks, personal experiences with English and Spanish, and perceived notions of prestige may be more important than sociolinguistic generation, at least for speakers not on the highest/lowest ends of adoption of or resistance to contact forms (see Figure 4). Social networks in particular have been found to be pivotal in the adoption of new linguistic forms

(O’Rourke and Potowski 2016; Dodsworth and Benton 2017; Carter and Lynch 2015; Carter 2007; Michnowicz et al. 2023a), and future research should focus on mapping speakers’ social networks and personal and community motivations with a goal of understanding these patterns of change. While the present data cannot provide a concrete answer to the important question of speaker motivation, the insights afforded by the swarm analysis can provide researchers with a roadmap to identify areas for more fine-grained research in the future. Additionally, future research should include more in-depth analyses of the role of individual speaker choices in the formation of a new bilingual community by utilizing the panoramic analysis provided by the variable swarm as a foundation/indicator of where those choices are likely to be most meaningful.

Author Contributions: Conceptualization, J.M. and R.R.; methodology, J.M., R.R., S.C., S.O. and G.G.; software, J.M., R.R., S.C., S.O. and G.G.; validation, J.M. and R.R.; formal analysis, J.M., R.R., S.C., S.O. and G.G.; investigation, J.M., R.R., S.C., S.O. and G.G.; resources, J.M. and R.R.; data curation, J.M. and R.R.; writing—original draft preparation, J.M., R.R., S.C., S.O. and G.G.; writing—review and editing, J.M. and R.R.; visualization, J.M.; supervision, J.M. and R.R.; project administration, J.M. and R.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the College of Humanities and Social Sciences Faculty Development Research Fund at North Carolina State University.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of North Carolina State University (protocol code 1586, 08/12/2010).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy reasons.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Participant demographics.

Participant	Sex	Birth Year	Generation	Region	Years in U.S.
2010-16	F	1957	G1	CAm	25
2014-13	M	1985	G1	CAm	19
2012-01	F	1957	G1	Mex	33
2013-21	F	1983	G1	Mex	1
2011-05	F	1988	G1	Mex	10
2013-03	M	1993	G1	Mex	10
2012-09	F	1992	G1	Mex	9
2013-15	M	1962	G1	Mex	35
2013-24	M	1984	G1	Mex	2
2013-23	M	1981	G1	Mex	2
2011-18	M	1992	G1	Mex	3
2013-26	M	1982	G1	Mex	1
2015-08	F	1992	G2	Mex	20
2011-21	F	1989	G2	CAm	Since birth
2014-10	F	1992	G2	CAm	Since birth
2012-15	F	1988	G2	CAm	Since birth
2012-22	F	1992	G2	CAm	Since birth
2012-08	F	1989	G2	CAm	Since birth
2013-19	M	1989	G2	Mex	Since birth
2013-18	F	1993	G2	Mex	Since birth
2014-19	F	1992	G2	Mex	Since birth
2013-28	M	1991	G2	Mex	Since birth
2013-02	M	1992	G2	Mex	Since birth

Appendix B

Table A2. Mexican G1. Results of the Multivariate One-Level Mixed-Effects Regression Model, speaker as a random factor. * Significant factors with a p -value < 0.05.

Factor		Log Odds	N	% Overt Pro	p-Value		
<i>Person/Number and Definiteness</i>					<0.001 *		
él-ella.definite		3.659	92	38.0			
yo		2.971	502	24.7			
ellos-ellas.definite		2.594	50	20.0			
tú.definite		1.860	30	13.3			
nosotros-nosotras		1.773	157	10.2			
tú.indefinite		0.883	104	3.8			
ellos-ellas.indefinite		−13.740	46	0.0			
<i>Switch Reference</i>					<0.001 *		
Complete switch		0.693	421	24.5			
No switch		−0.210	480	16.0			
Partial switch		−0.484	80	16.2			
<i>Lexical Content</i>					0.0047 *		
Estimative		0.916	46	52.2			
Stative		−0.164	221	23.1			
External activity		−0.358	555	16.2			
Mental activity		−0.393	159	17.6			
<i>Distinctiveness of TAM</i>					0.013 *		
Non-distinctive		0.27	176	26.7			
Distinctive		−0.27	805	18.1			
<i>Reflexivity</i>					0.0185 *		
Non-reflexive		0.36	865	20.7			
Reflexive		−0.36	116	12.1			
n	df	intercept	overall proportion	AIC	R2.fixed	R2.random	R2.total
981	15	−4.187	0.197	859.63	0.781	0.012	0.793

Table A3. Central American G1. Results of the Multivariate One-Level Mixed-Effects Regression Model, speaker as a random factor. * Significant factors with a p -value < 0.05.

Factor			Log Odds	N	% Overt Pro	p-Value	
<i>Person/Number and Definiteness</i>						<0.001 *	
él-ella.definite			8.721	25	56.0		
yo			7.637	107	20.6		
ellos-ellas.definite			6.251	21	9.5		
nosotros-nosotras			5.868	25	12		
tú.definite			−8.910	3	0.0		
ellos-ellas.indefinite			−9.191	10	0.0		
tú.indefinite			−10.376	1	0.0		
<i>Switch Reference</i>						<0.001 *	
Partial switch			0.613	17	35.3		
Complete switch			0.567	87	28.7		
No switch			−1.180	88	11.4		
<i>Lexical Content</i>						0.009 *	
Stative			5.187	62	37.1		
External activity			4.158	98	16.3		
Mental activity			3.310	28	7.1		
Estimative			−12.655	4	0.0		
<i>Distinctiveness of TAM</i>						0.849	
Non-distinctive			0.049	53	26.4		
Distinctive			−0.049	139	119.4		
<i>Reflexivity</i>						0.144	
Non-reflexive			0.463	158	23.4		
Reflexive			−0.463	34	11.8		
n	df	intercept	overall proportion	AIC	R2.fixed	R2.random	R2.total
192	15	−13.359	0.214	173.926	0.88	0.0	0.88

Table A4. Mexican G2. Results of the Multivariate One-Level Mixed-Effects Regression Model, speaker as a random factor. * Significant factors with a p -value < 0.05.

Factor				Log Odds	N	% Overt Pro	p-Value
<i>Person/Number and Definiteness</i>							<0.001 *
él-ella.definite				5.596	13	15.4	
yo				5.137	396	17.9	
ellos-ellas.definite				5.076	60	15.0	
ellos-ellas.indefinite				4.999	12	8.3	
tú.indefinite				3.910	41	4.9	
tú.definite				−11.768	24	0.0	
nosotros-nosotras				−12.951	52	0.0	
<i>Distinctiveness of TAM</i>							<0.001 *
Non-distinctive				0.564	106	20.8	
Distinctive				−0.564	492	12.8	
<i>Lexical Content</i>							0.00144 *
Estimative				1.269	34	50.0	
Mental activity				−0.247	125	15.2	
External activity				−0.348	310	10.6	
Stative				−0.674	129	12.4	
<i>Reflexivity</i>							0.00163 *
Non-reflexive				1.151	535	15.7	
Reflexive				−1.151	63	1.6	
<i>Switch Reference</i>							0.0431 *
Complete switch				0.646	236	18.2	
No switch				0.101	315	12.7	
Partial switch				−0.747	47	4.3	
n	df	intercept	overall proportion	AIC	R2.fixed	R2.random	R2.total
598	15	−7.662	0.142	420.84	0.905	0.013	0.918

Table A5. Central American G2. Results of the Multivariate One-Level Mixed-Effects Regression Model, speaker as a random factor. * Significant factors with a p -value < 0.05.

Factor			Log Odds	N	% Overt Pro	p-Value	
Person/Number and Definiteness						<0.001 *	
tú.definite			10.809	1	100.0		
tú.indefinite			2.161	5	80.0		
él-ella.definite			−0.570	137	37.2		
ellos-ellas.definite			−2.268	78	21.8		
yo			−2.424	209	21.5		
ellos-ellas.indefinite			−3.632	6	16.7		
nosotros-nosotras			−4.077	58	10.3		
Switch Reference						<0.001 *	
Complete switch			0.885	201	34.8		
Partial switch			−0.127	37	32.4		
No switch			−0.757	256	16.8		
Lexical Content						0.481	
External activity			0.404	266	28.2		
Stative			0.162	150	22.0		
Mental activity			0.053	63	23.8		
Estimative			−0.619	15	13.3		
Distinctiveness of TAM						0.713	
Non-distinctive			0.05	140	29.3		
Distinctive			−0.05	354	23.7		
Reflexivity						0.428	
Non-reflexive			0.153	439	25.7		
Reflexive			−0.153	55	21.8		
n	df	intercept	overall proportion	AIC	R2.fixed	R2.random	R2.total
494	15	0.308	0.253	485.02	0.322	0.133	0.455

Notes

- ¹ Note that many lexical FPs are also analyzable as discourse markers.
- ² It should be noted that in our own data, the majority of the auditorily labiodental tokens were also approximants rather than fricatives. Thus, the presence of the labiodental approximant [ʋ] complicates the classification of potential contact forms, as a bilingual speaker could produce a hybrid form that utilizes the labiodental point of articulation from English alongside the approximant mode of articulation from Spanish. In this way, bilingual labiodentals may not correspond exactly with English fricative /v/.
- ³ Technical aspects regarding FASE are described in greater depth in Wilbanks (2015) and Ronquest et al. (2020). Regarding the reliability of automatic alignment with FASE, Wilbanks (2015) compares FASE alignment to the alignment produced by trained human phoneticians. He finds that the refined, adapted FASE segmentation was similar to human segmentation: boundary differences between the two human coders had a mean of 14.47 ms, compared to a mean difference of 20.81 ms between the human coders and the trained FASE model.
- ⁴ For example, the full model AIC was compared to the AIC of a model with all variables except one (e.g., Switch Reference). The difference in AIC values indicates the importance of the variable in the full model, with larger differences denoting a stronger effect on the model. Kapatsinski (2012) demonstrates why this method is superior to the range of coefficients, the traditional method of determining the constraint hierarchy in sociolinguistic studies, which can be biased towards variables/factors with more levels.
- ⁵ An acoustic analysis of phonological filled pauses, as well as an examination of lexical pauses (e.g., *sea, este*) and silent pauses is forthcoming.
- ⁶ Data are frequently skewed to the right when there is a lower boundary to the measurement; in this case, COG cannot be a negative number. See <https://www.itl.nist.gov/div898/handbook/eda/section3/histogr6.htm>, accessed on 4 March 2023.
- ⁷ There are only two Central American G1 speakers in the present corpus, and some of these variables may achieve significance in a larger sample. Still, the match between Mexican and Central American G1 speakers speaks to the robustness of these constraints across Spanish varieties, even when few speakers are analyzed.
- ⁸ However, in some contexts, bilinguals have been found to produce fewer FPs than monolinguals but higher rates of silent pauses, which were not studied here (García-Amaya 2022). Additionally, although not coded in our data, impressionistically some monolinguals may have compensated for lower rates of phonological FPs by using more lexical FPs (*pues, este*, etc.). Further study is warranted and underway.
- ⁹ A more detailed analysis of and <v> in NC Spanish is forthcoming.
- ¹⁰ <http://www.sthda.com/english/wiki/correlation-matrix-an-r-function-to-do-all-you-need>, accessed on 4 March 2023.

References

- Abreu, Laurel. 2012. Subject pronoun expression and priming effects among bilingual speakers of Puerto Rican Spanish. In *Selected Proceedings of the 14th Hispanic Linguistics Symposium*. Edited by Kimberly Geeslin and Manuel Díaz-Campos. Somerville: Cascadilla Proceedings Project, pp. 1–8.
- Alvord, Scott, and Brandon Rogers. 2014. Miami-Cuban Spanish vowels in contact. *Sociolinguistic Studies* 8: 139–70. [CrossRef]
- Anderson, Tyker Kimball, and Almeida Jacqueline Toribio. 2007. Attitudes towards lexical borrowing and intra-sentential code-switching among Spanish-English bilinguals. *Spanish in Context* 4: 217–40. [CrossRef]
- Aaron, Jessi Elana. 2004. The Gendered Use of *salirse* in Mexican Spanish: Si me salía yo con las amigas, se enojaba. *Language in Society* 33: 585–607. [CrossRef]
- Bayley, Robert, and Lucinda Pease-Alvarez. 1997. Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9: 349–71. [CrossRef]
- Bayley, Robert, Norma L. Cárdenas, Belinda Treviño Schouten, and Carlos Martín Vélez Salas. 2012. Spanish dialect contact in San Antonio, Texas: An exploratory study. In *Selected Proceedings of the 14th Hispanic Linguistics Symposium*. Edited by Kimberly Geeslin and Manuel Díaz-Campos. Somerville: Cascadilla Proceedings Project, pp. 48–60.
- Boersma, Paul, and David Weenink. 2022. *Praat: Doing Phonetics by Computer [Computer Program]*, Version 6.3.02. Available online: <https://www.praat.org/> (accessed on 29 November 2022).
- Boomershine, Amanda, and Rebecca Ronquest. 2019. Teaching pronunciation to Spanish heritage speakers. In *Key Issues in the Teaching of Spanish Pronunciation*. Edited by Rajiv Rao. London: Routledge, pp. 288–303.
- Britain, David, and Peter Trudgill. 1999. Migration, new-dialect formation and sociolinguistic refunctionalisation: Reallocation as an outcome of dialect contact. *Transactions of the Philological Society* 97: 245–56. [CrossRef]
- Carolina Demography. 2021. North Carolina's Hispanic Community: Snapshot (2021). Available online: <https://www.ncdemography.org/> (accessed on 24 February 2023).
- Carter, Phillip M. 2005. Prosodic variation in SLA: Rhythm in an urban North Carolina Hispanic community. *University of Pennsylvania Working Papers in Linguistics* 11: 59–71.
- Carter, Phillip M. 2007. Phonetic Variation and Speaker Agency: Mexicana Identity in a North Carolina Middle School. *Penn Working Papers in Linguistics* 13: 1–15.

- Carter, Phillip M. 2013. Shared spaces, shared structures: Latino social formation and African American English in the U.S. South. *Journal of Sociolinguistics* 17: 66–92. [CrossRef]
- Carter, Phillip M., and Andrew Lynch. 2015. Multilingual Miami: Current trends in sociolinguistic research. *Language and Linguistics Compass* 9: 369–85. [CrossRef]
- Carter, Phillip M., and Tonya Wolford. 2016. Cross-generational prosodic convergence in South Texas Spanish. *Spanish in Context* 13: 29–52. [CrossRef]
- Chetty, Sarah. 2018. To /b/ or not to /b/: On the production of the graphemes <bv> in heritage Spanish [Poster presentation]. In *The 2018 Graduate Research Symposium*. Raleigh: NC State University. Available online: <https://grad.ncsu.edu/wp-content/uploads/2018/04/sarahchettyposter.pdf> (accessed on 24 February 2023).
- Chládková, Kateřina, Paola Escudero, and Paul Boersma. 2011. Context-specific acoustic differences between Peruvian and Iberian Spanish vowels. *JASA* 130: 416–28. [CrossRef]
- Cobb, Katherine, and Miquel Simonet. 2015. Adult Second Language Learning of Spanish Vowels. *Hispania* 98: 47–60. [CrossRef]
- Colantoni, Laura, and Irina I. Marinescu. 2010. The scope of stop weakening in Argentine Spanish. In *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*. Edited by Marta Ortega-Llebaria. Somerville: Cascadia Proceedings Project, pp. 100–14.
- Coloma, Germán. 2012. *The Importance of Ten Phonetic Characteristics to Define Dialect Areas in Spanish*. Dialectologia: Revista Electrónica, pp. 1–26.
- Dodsworth, Robin, and Richard A. Benton. 2017. Social network cohesion and the retreat from Southern vowels in Raleigh. *Language in Society* 46: 371–405. [CrossRef]
- Drager, Katie, and Jennifer Hay. 2012. Exploiting random intercepts: Two case studies in sociophonetics. *Language Variation and Change* 24: 59–78. [CrossRef]
- Eckert, Penelope. 2005. *Variation, Convention, and Social Meaning [Plenary Address]*. Oakland: The Annual Meeting of the Linguistic Society of America. Available online: <http://www.justinecassell.com/discourse09/readings/EckertLSA2005.pdf> (accessed on 19 May 2023).
- Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41: 87–100. [CrossRef]
- Erker, Daniel. 2017. The limits of named language varieties and the role of social salience in dialectal contact: The case of Spanish in the United States. *Language and Linguistics Compass* 11: 1–20. [CrossRef]
- Erker, Daniel, and Joana Bruso. 2017. Uh, bueno, em . . . : Filled pauses as a site of contact-induced change in Boston Spanish. *Language Variation and Change* 29: 205–44. [CrossRef]
- Escobar, Anna María, and Kim Potowski. 2015. *El español de los Estados Unidos*. Cambridge: Cambridge University Press.
- File-Muriel, Richard J., and Earl K. Brown. 2011. The gradient nature of s-lenition in Caleño Spanish. *Language Variation and Change* 23: 223–43. [CrossRef]
- García-Amaya, Lorenzo. 2009. New findings on fluency measures across three different learning contexts. In *Selected Proceedings of the 11th Hispanic Linguistics Symposium*. Edited by Joseph Collentine, Maryellen García, Barbara Lafford and Francisco Marcos Marín. Somerville: Cascadia Press, pp. 68–80.
- García-Amaya, Lorenzo. 2022. An investigation into utterance-fluency patterns of advanced LL bilinguals: Afrikaans and Spanish in Patagonia. *Linguistic Approaches to Bilingualism* 12: 163–90. [CrossRef]
- Grabe, Esther, and Ee Ling Low. 2002. Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology* 7: 515–46.
- Hernández, José Esteban. 2002. *Accommodation in a Dialect Contact Situation*. San José: Revista de filología y lingüística de la Universidad de Costa Rica, pp. 93–110.
- Howe, Chad, and Philip Limerick. 2020. Perceptions of Spanish among Spanish Heritage Speakers in the Southeastern United States through Computer-Mediated Communication. In *Spanish in the United States and across Domains*. Edited by Francisco Salgado-Robles and Edwin Lamboy. Leiden: Brill, pp. 364–87.
- Hualde, José Ignacio. 2005. *The Sounds of Spanish*. Cambridge: Cambridge University Press.
- Hualde, José Ignacio, Miquel Simonet, and Marianna Nadeu. 2011. Consonant lenition and phonological recategorization. *Laboratory Phonology* 2: 301–29. [CrossRef]
- Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3: 359–83. [CrossRef]
- Kapatsinski, Vsevolod. 2012. *Towards a de-Ranged Study of Variation [Conference Presentation]*. Washington, DC: Georgetown University Round Table on Linguistics, Measured Language: Quantitative Approaches to Acquisition, Assessment, Processing, and Variation. Available online: <http://pages.uoregon.edu/vkapatsi/DerangedPaper.pdf> (accessed on 24 February 2023).
- Kerswill, Paul. 2013. Koineization. In *The Handbook of Language Variation and Change*. Edited by J. K. Chambers and Natalie Schilling. Oxford: John Wiley & Sons.
- Knightly, Leah, SunAh Jun, Janet Oh, and Terry Kit-fong Au. 2003. Production benefits of childhood overhearing. *JASA* 114: 465–74. [CrossRef] [PubMed]
- Knouse, Stephanie M., Renee Neves, Erk Ortiz, and Daria Acosta-Rua. 2022. “Le está haciendo un disservice”: Overt Attitudes toward Language Contact Phenomena in the Upstate of South Carolina. *Hispania* 105: 173–94. [CrossRef]

- Labov, William. 1963. The social motivation of a sound change. *Word* 19: 273–309. [CrossRef]
- Labov, William, Ingrid Rosenfelder, and Josef Fruehwald. 2013. One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language* 89: 30–65. [CrossRef]
- Ladefoged, Peter, and Sandra Disner. 2012. *Vowels and Consonants*. Chichester: John Wiley and Sons.
- Limerick, Pilip P. 2019. The discursive distribution of subject pronouns in Spanish spoken in Georgia: A weakening of pragmatic constraints? *Studies in Hispanic and Lusophone Linguistics* 12: 97–126. [CrossRef]
- Limerick, Philip P. 2021. First-person plural subject pronoun expression in Mexican Spanish spoken in Georgia. *Studies in Hispanic and Lusophone Linguistics* 14: 411–32. [CrossRef]
- Lipski, John M. 1994. *Latin American Spanish*. London: Longman.
- Lipski, John M. 2005. Code-switching or borrowing? No sé so no puedo decir, you know. In *Selected Proceedings of the Second Workshop on Spanish Sociolinguistics*. Edited by Lotfi Sayahi and Maurice Westmoreland. Somerville: Cascadilla Proceedings Project, pp. 1–15.
- Lipski, John M. 2008. *Varieties of Spanish in the United States*. Washington, DC: Georgetown University Press.
- Lipski, John M. 2014. Spanish-English code-switching among low-fluency bilinguals: Towards an expanded typology. *Sociolinguistic Studies* 8: 23. [CrossRef]
- Lipski, John M. 2020. Equatorial Guinea Spanish non-continuant /d/: More than a generic L2 trait. In *Spanish Phonetics and Phonology in Contact: Studies from Africa, the Americas, and Spain*. Edited by Rajiv Rao. Amsterdam: John Benjamins, pp. 13–32.
- Low, Ee-Ling, and Esther Grabe. 1995. Prosodic patterns in Singapore English. In *Proceedings of the International Congress of Phonetic Sciences, Stockholm*. Edited by Kjell Elenius and Peter Branderud. Stockholm: KTH and Stockholm University Press, pp. 636–39.
- Mariano, Paolo. 2014. Correlatore [Computer Program]. Available online: http://www.lfsag.unito.it/correlatore/readme_en.html (accessed on 14 September 2022).
- Mata, Rodolfo. 2022. Bilingualism is good but codeswitching is bad: Attitudes about Spanish in contact with English in the Tijuana-San Diego border area. *Critical Inquiry in Language Studies* 1–22. [CrossRef]
- Mendoza-Denton, Norma. 2002. Language and Identity. In *The Handbook of Language Variation and Change*. Edited by J. K. Chambers, Peter Trudgill and Natalie Schilling-Estes. Oxford: Blackwell, pp. 475–99.
- Mendoza-Denton, Norma. 2010. Individuals and communities. In *The Sage Handbook of Sociolinguistics*. Edited by Ruth B. Wodak, Barbara Johnstone and Peter Kerswill. Los Angeles: Sage Publications, pp. 181–91.
- Menke, Mandy R., and Timothy L. Face. 2010. Second language Spanish vowel production: An acoustic analysis. *Studies in Hispanic and Lusophone Linguistics* 3: 181–214. [CrossRef]
- Michnowicz, Jim. 2015. Subject Pronoun Expression in Contact with Maya in Yucatan Spanish. In *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective*. Edited by Ana Carvalho, Rafael Orozco and Naomi Lapidus Shin. Washington, DC: Georgetown University Press, pp. 103–22.
- Michnowicz, Jim, Alex Hyler, James Shepherd, and Sonya Trawick. 2018. Spanish in North Carolina: English-origin loanwords in a newly forming Hispanic community. In *Language Diversity in the New South*. Edited by Jeffrey Reaser, Eric Wilbanks, Walt Wolfram and Karissa Wojcik. Chapel Hill: University of North Carolina Press, pp. 289–305.
- Michnowicz, Jim, Rebecca Ronquest, Bailey Ambrister, Hannah Bain, Nick Chisholm, Rebecca Greene, Lindsey Bull, and Anne Elkins. 2023a. Perceptions of Inclusive Language in the Spanish of the Southeast: Data from a Large Classroom Project. *Spanish in Context*. Available online: <https://www.jbe-platform.com/content/journals/10.1075/sic.00084.mic> (accessed on 6 July 2023).
- Michnowicz, Jim, Sonya Trawick, and Ronquest Ronquest. 2023b. Spanish language maintenance and shift in a newly-forming community in the southeastern United States: Insights from a large-class survey. *Hispanic Studies Review* 7. Available online: <https://hispanicstudiesreview.cofc.edu/article/77601-spanish-language-maintenance-and-shift-in-a-newly-forming-community-in-the-southeastern-united-states-insights-from-a-large-class-survey> (accessed on 7 July 2023).
- Montes-Alcalá, Cecilia, and Lindsey Sweetnich. 2014. Español en el sureste de EE. UU.: El papel de las actitudes lingüísticas en el mantenimiento o pérdida de la lengua. *Revista Internacional de Lingüística Iberoamericana* 23: 77–92. [CrossRef]
- Myers-Scotton, Carol. 1995. *Social Motivations for Codeswitching: Evidence from Africa*. Oxford: Oxford University Press.
- Navarro Tomás, Tomás. 1918. *Manual de Pronunciación Española*, 12th ed. Madrid: CSIC.
- O'Rourke, Erin, and Kim Potowski. 2016. Phonetic accommodation in a situation of Spanish dialect contact: Coda/s/and/r̄/in Chicago. *Studies in Hispanic and Lusophone Linguistics* 9: 355–99. [CrossRef]
- Otheguy, Ricardo, Ana Celia Zentella, and David Livert. 2007. Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language* 83: 770–802. [CrossRef]
- Otheguy, Ricardo, and Ana Celia Zentella. 2012. *Spanish in New York: Language Contact, Dialectal Leveling, and Structural Continuity*. Oxford: Oxford University Press.
- Pew Research Center. 2014. Demographic Profile of Hispanics in North Carolina. Available online: <http://www.pewhispanic.org/states/state/nc> (accessed on 2 October 2022).
- Poplack, Shana. 1978. Dialect acquisition among Puerto Rican bilinguals. *Language in Society* 7: 89–103. [CrossRef]
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18: 581–618. [CrossRef]
- Poplack, Shana. 1988. Contrasting patterns of code-switching in two communities. *Codeswitching: Anthropological and sociolinguistic Perspectives* 48: 215–44.

- Quilis, Antonio, and Manuel Esgueva. 1983. Realización de los fonemas vocálicos españoles en posición fonética normal. In *Estudios de fonética*. Edited by Manuel Esgueva and Margarita Cantarero. Madrid: Consejo Superior de Investigaciones Científicas, pp. 159–251.
- Rangel, Natalie, Verónica Loureiro-Rodríguez, and María Irebe Moyna. 2015. “Is that what I sound like when I speak?”: Attitudes towards Spanish, English, and code-switching in two Texas border towns. *Spanish in Context* 12: 177–98. [CrossRef]
- Rao, Rajiv. 2014. On the status of the phoneme /b/ in heritage speakers of Spanish. *Sintagma* 26: 37–54.
- Rao, Rajiv. 2015. Manifestations of /bdg/ in heritage speakers of Spanish. *Heritage Language Journal* 12: 48–74. [CrossRef]
- Rao, Rajiv, and Rebecca Ronquest. 2015. The heritage Spanish phonetic/phonological system: Looking back and moving forward. *Studies in Hispanic and Lusophone Linguistics* 8: 403–14. [CrossRef]
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: <https://www.R-project.org/> (accessed on 3 October 2022).
- Ronquest, Rebecca. 2012. An Acoustic Analysis of Heritage Spanish Vowels. Ph.D. thesis, Indiana University, Bloomington, Indiana.
- Ronquest, Rebecca, Jim Michnowicz, Eric Wilbanks, and Cortés Claudia. 2020. Examining the (mini-)variable swarm in the Spanish of the Southeast. In *Hispanic Linguistics: Current Issues and New Directions*. Edited by Alfonso Morales-Front, Michael Ferreira, Ronald Leow and Cristina Sanz. Amsterdam: John Benjamins, pp. 303–25.
- Shin, Naomi Lapidus. 2013. Women as leaders of language change: A qualification from the bilingual perspective. In *Selected Proceedings of the 6th Workshop on Spanish Sociolinguistics*. Edited by Ana M. Carvalho and Sara Beaudrie. Somerville: Cascadia Proceedings Project, pp. 135–47.
- Shin, Naomi Lapidus, and Ricardo Otheguy. 2009. Shifting sensitivity to continuity of reference: Subject pronoun use in Spanish. In *Español en Estados Unidos y en otros contextos: Cuestiones sociolingüísticas, políticas, y pedagógicas*. Edited by Manel Lacorte and Jennifer Leeman. Madrid: Iberoamericana/Vervuert Verlag, pp. 111–36.
- Shin, Naomi Lapidus, and Ricardo Otheguy. 2013. Social class and gender impacting change in bilingual settings: Spanish subject pronoun use in New York. *Language in Society* 42: 429–52. [CrossRef]
- Shousterman, Cara. 2014. Speaking English in Spanish Harlem: The role of rhythm. *University of Pennsylvania Working Papers in Linguistics* 20: 18.
- Silva-Corvalán, Carmen. 1994. The gradual loss of mood distinctions in Los Angeles Spanish. *Language Variation and Change* 6: 255–72. [CrossRef]
- Silva-Corvalán, Carmen. 2008. The limits of convergence in language contact. *Journal of Language Contact* 2: 213–24. [CrossRef]
- Sorace, Antonella. 2004. Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism Language and Cognition* 7: 143–5. [CrossRef]
- Sorace, Antonella. 2005. Selective Optionality in Language Development. In *Syntax and Variation: Reconciling the Biological and the Social*. Edited by Leonie Cornips and Karen Corrigan. Amsterdam: John Benjamins, pp. 55–80.
- Tagliamonte, Sali A. 2011. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Hoboken: John Wiley and Sons.
- Thomas, Erik. 2015. What a swarm of variables tells us about the formation of Mexican American English [Conference presentation]. Paper presented at LAVIS IV, Raleigh, NC, USA, August 28. Available online: https://www.academia.edu/15228944/What_a_Swarm_of_Variables_Tells_Us_about_the_Formation_of_Mexican_American_English (accessed on 1 November 2022).
- Toribio, Almeida Jacqueline. 2004. Spanish/English speech practices: Bringing chaos to order. *International Journal of Bilingual Education and Bilingualism* 7: 133–54. [CrossRef]
- Toribio, Almeida Jacqueline. 2011. Code-Switching among US Latinos. In *The Handbook of Hispanic Sociolinguistics*. Hoboken: Wiley, pp. 530–52.
- Torres, Lourdes. 2002. Bilingual discourse markers in Puerto Rican Spanish. *Language in Society* 31: 65–83. [CrossRef]
- Torres, Lourdes. 2011. Spanish in the United States: Bilingual discourse markers. In *The Handbook of Hispanic Sociolinguistics*. Edited by Manuel Díaz-Campos. Malden: Wiley-Blackwell, pp. 491–503.
- Torres, Lourdes, and Kim Potowski. 2008. A comparative study of bilingual discourse markers in Chicago Mexican, Puerto Rican, and MexiRican Spanish. *International Journal of Bilingualism* 12: 263–79. [CrossRef]
- Trovato, Adriano. 2017. A Sociophonetic Analysis of Contact Spanish in the United States: Labiodentalization and Labial Consonant Variation. Ph.D. thesis, University of Texas, Austin, TX, USA.
- Trudgill, Peter. 1986. *Dialects in Contact*. Oxford: Blackwell.
- U.S. Census Bureau. 2020. Hispanic or Latino Origin by Specific Origin. 2016–20 American Community Survey 5-Year Estimates. Available online: <https://data.census.gov/cedsci/table?q=Latino%20Origin%20NC&tid=ACSDT5Y2020.B03001> (accessed on 31 October 2022).
- Wei, Taiyun, and Viliam Simko. 2021. R Package ‘Corrplot’: Visualization of a Correlation Matrix (Version 0.92). Available online: <https://github.com/taiyun/corrplot> (accessed on 5 November 2022).
- Wilbanks, Eric. 2015. The Development of FASE (Forced Alignment System for Español) and Implications for sociolinguistic research. Paper presented at New Ways of Analyzing Variation 44, Toronto, ON, Canada, October 22–25. Available online: https://ericwilbanks.github.io/files/wilbanks_nwav_2015.pdf (accessed on 20 May 2023).
- Willis, Erik. 2005. An Initial Examination of Southwest Spanish Vowels. *Southwest Journal of Linguistics* 24: 185–98.

- Wolfram, Walt, Mary Kohn, and Erin Callahan-Price. 2011. Southern-Bred Hispanic English: An Emerging Socioethnic Variety. In *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics*. Edited by Jim Michnowicz and Robin Dodsworth. Somerville: Cascadia Proceedings Project, pp. 1–13.
- Wolfram, Walt, Phillip Carter, and Beckie Morello. 2004. Emerging Hispanic English: New dialect formation in the American South. *Journal of Sociolinguistics* 8: 339–58. [[CrossRef](#)]
- Woolard, Kathryn A. 1999. Strategies of simultaneity and bivalency in bilingual communication. *Journal of Linguistic Anthropology* 8: 3–29. [[CrossRef](#)]
- Zentella, Ana Celia. 1990. Lexical leveling in four New York City Spanish dialects: Linguistic and social factors. *Hispania* 73: 1094–105. [[CrossRef](#)]
- Zentella, Ana Celia. 1997. *Growing up Bilingual: Puerto Rican Children in New York*. Malden: Blackwell Publishers.
- Zúñiga, Victor, and Rubén Hernández-León, eds. 2005. *New Destinations: Mexican Immigration in the United States: Community Formation, Local Responses and Inter-Group Relations*. New York: Russell Sage Foundation.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.