

Aspectual Differences in the Preterite and Imperfect in US Spanish: An Analysis of Their
Use by Three Generation English Spanish Bilinguals

by

Leslie Del Carpio

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved August 2022 by the
Graduate Supervisory Committee:

Sara Beaudrie, Co-Chair
Alvaro Cerrón-Palomino, Co-Chair
Michael Gradoville
Ana Maria Carvalho

ARIZONA STATE UNIVERSITY

December 2022

ABSTRACT

As critical approaches to Heritage Language (HL) instruction are increasingly more widespread, it is now more pertinent than ever to conduct research on Spanish linguistic variation that reveals systematicity and refutes hegemonic notions of ‘incorrectness’. This variationist study examines the use and distribution of the Preterite and Imperfect past tenses in Spanish. The study analyzes 30 bilingual English Spanish speakers who reside in southern and central Arizona by using sociolinguistic interviews from two Arizona corpora. These data provided by these interviews was analyzed using the Rbrul and Rstudio software. The linguistic factors analyzed were aspectual interpretation, clause type, grammatical person and number, verb lemma and verb frequency. By the same token, the extra linguistic factors analyzed were generation, language dominance, age and sex. The findings in this study reveal distribution of the Preterite and Imperfect in the data revealed that both forms were used at nearly equal rates with the Preterite (53.7%) being used slightly more than the Imperfect (46.3%).

The results of the quantitative analysis revealed that in order of magnitude, aspectual function, generation, and frequency of the verb were the predictors that favored the Preterite and the Imperfect. While the majority of Preterite and Imperfect uses adhered to the canonical uses of these forms, an interaction between generation and aspectual function showed significance when the Preterite is used with a habitual aspectual function by both the second and third generation. These results show that the Preterite and Imperfect carry a degree of variation that goes beyond the traditional understandings of these forms. Lastly, the results of this study emphasize the need for additional research that aids in the understanding of the characteristics of US Spanish to

dispel misconceptions about the Spanish spoken in the U.S by all, especially HL speakers. It is only by understanding the evolution of the Spanish in the US that researchers and instructors can contribute to a critical language awareness in HL instruction that revindicates their ways of speaking.

To my family,

ACKNOWLEDGMENTS

As a student of ASU, I acknowledge that the Tempe campus sits on the ancestral homelands of those American Indian tribes that have inhabited this place for centuries, including the Akimel O’odham (Pima) and Pee Posh (Maricopa) peoples. Thank you.

I would like to express my deep gratitude to my dissertation committee members Dr. Sara Beaudrie, Dr. Alvaro Cerrón-Palomino, Dr. Michael Gradoville and Dr. Ana Maria Carvalho for their patience, professional guidance, and encouragement. I am extremely grateful for all the times each of you met with me and provided me with your wisdom and guidance. I would also like to express my gratitude to Dr. Anna María Escobar who has always been there to provide me with continuous ongoing support.

I am also grateful to the friends I met at SILC who became my family during my time at ASU. To my student mentor, Rosti for always answering any question I had along the way. To Anna, Anne, Beto, Tomás, Valeria, among others; thank you for all the good times, academic and non-academic conversations, and advice and for being there during the good and not so good times. Thank you as well to Keren and Paola for all the times they gladly shared their input on any question I had. I would also like to express a big thank you to Dr. Joel Judd who help me through the last steps of the dissertation.

To my grandma Isabel, my mom Gabriela, my sister Kiara and my brothers Brandon and Mosiah, thank you for being part of my family, thank you for providing me with support and for brightening my day one way or another and for always believing in me. To my loving dad Walter, words cannot explain how much I love you and miss you. Lastly, to Apollo and Milo, you were the best company I could have asked for these past four years. I couldn’t imagine having gone through this process without both of you 🐾.

TABLE OF CONTENTS

	Page
LIST OF TABLES	ix
LIST OF FIGURES.....	x
TRANSCRIPTION CONVENTIONS	xi
CHAPTER	
INTRODUCTION	1
The Spanish Past Tense	3
Aspectual Domains of the Spanish Past Tense	4
Aspectual Domains of the English Past Tense	7
Languages in Contact	9
Spanish Speakers in Arizona	10
Heritage Speakers of Spanish and Heritage Language Learners	14
Rationale	15
Organization of the Dissertation	17
LITERATURE REVIEW	19
Introduction	19
Morphosyntactic Variables.....	20
The Preterite and Imperfect in Monolingual Spanish Speakers	22
The Preterite and Imperfect in Bilingual Speakers in the US.....	27
Cross-generational studies	35
Summary	37

CHAPTER	Page
METHODOLOGY	39
Introduction	39
The Rationale	39
Research Questions	41
The Data	41
Envelope of Variation	46
Factorr Groups	48
Internal Predictors	48
External Predictors	54
Analysis	59
Summary	60
RESULTS AND DISCUSSION	62
Introduction.....	62
Overall Preterite Versus Imperfect Use	63
The Preterite.....	68
The Imperfect.....	72
Statistical Modeling	77
Preterite and Imperfect Uses by Generation	79
Generation and Aspectual Function	80
Generation and Verb Frequency.....	84

CHAPTER	Page
The Preterite and Imperfect: Past and Present	87
The Current Study and Studies on Monolingual Speakers	88
The Current Study and Studies on Bilingual Speakers	90
Spanish in the US	92
Pedagogical Implications	94
Summary	99
 CONCLUSION	 101
Introduction	101
Summary	102
Contributions to the Field of Language Variation Research	103
Contributions to the Field of US Spanish	105
Contributions to the Field of Spanish As a Heritage Language	107
Limitations	109
Future Directions	110
 REFERENCES	 114
 APPENDIX	
A THE TWELVE BILINGUAL DOMINANCE SCALE QUESTIONS AND THE SCORING PROCEDURE BY DUNN & TREE (2009).....	123
B BILINGUAL LANGUAGE PROFILE: SPANISH ENGLISH COMPETENCE BY CARVALHO (2012-).....	126
C VERB FREQUENCY AND PROPORTION OF PRETERITE USE	128

LIST OF TABLES

Table	Page
1. Participants Background Information	44
2. Overall Frequencies of the Preterite and Imperfect	63
3. Frequencies of the Imperfect and Preterite and Internal Predictors	64
4. Multivariate Analysis in Rbrul	78
5. Distribution of Preterite and Imperfect Per Generation	80
6. Fixed Effects of the Mixed-Effects Model Between Generation X Aspectual Function.....	83
7. Verbs with High Predictor Weight with Frequency and Proportion Values.....	85
8. Verbs Used in Preterite with Habitual Aspect by G2 and G3	86
9. Presenting Preterite and Imperfect Variation to HL learners.....	110

LIST OF FIGURES

Figure	Page
1. Aspectual Distributions (Comrie, 1976, pp.25)	5
2. Elongation and Filler Rated by Bilingual Dominance Scale (Dunn & Fox Tree, 2009, pp.286)	57
3. Conditional Tree of the Preterite	70
4. Conditional Tree of the Imperfect	74
5. Proportion of the Preterite Versus the Imperfect with Respective Aspectual Categories	81

TRANSCRIPTION CONVENTIONS

/.../	Interviewer Interruption
XY	Anonymization of Identifiable Information
“ ”	Direct/Quoted Speech
XXX	Unintelligible Speech

All other punctuation marks (periods, commas, question marks, exclamation points) are used as in “standard” writing)

CHAPTER 1

INTRODUCTION

The use of the Spanish past tense and its aspectual features, especially the use of the Imperfect and the Preterite¹, has captivated the attention of many scholars in the field of Spanish linguistics. In particular, researchers in the area of Second Language Acquisition (SLA) have undertaken the study of this linguistic feature by focusing on the production and perception of the past tenses (e.g., Ayoun & Salaberry, 2005; Quesada, 2013; Slabakova, 2002; Montrul & Slabakova, 2000). The present dissertation takes a variationist approach to the study of the Preterite and the Imperfect use by three generation English Spanish bilinguals living in Arizona.

A variation approach was chosen for a few different reasons. (1) A variationist approach allows for a quantitative analysis of the forms being observed. In addition, this approach observes the role of internal as well as external factors and their relative effects through a multivariate analysis. (2) The research on the Preterite and Imperfect outside of experimental methodologies remains scarce. This type of research is pertinent as non-experimental data such as sociolinguistic interviews allow the opportunity to analyze more naturalistic speech. (3) To date, this is the first variationist approach on the Preterite and Imperfect spoken by Spanish English bilinguals. All the data analyzed in this

¹ Following Comrie (1976), I am using capital letters with the names of language-specific forms. Conversely, I am using lower case with typological semantic categories.

dissertation come from two Arizona corpora, CESA (Carvalho, 2012-)² and CEPA (Cerrón-Palomino, 2012-)³.

Scholars have often assumed that the Preterite and Imperfect occupy two different spaces of past tense reference in Spanish and that their uses are rarely interchangeable. However, this dissertation explores if the use of the Preterite and the Imperfect differ from what the traditional Spanish grammars describe. This initiative is motivated by previous works that have found variation in the use of the Preterite and Imperfect. These works are the seminal qualitative research on Spanish English bilinguals in Los Angeles and New York (Silva-Corvalán, 1994; Zentella, 1997), as well as more recent variationist findings in different monolingual Spanish varieties such as Argentinean and Puerto Rican Spanish (Delgado-Díaz, 2018, Delgado-Díaz, 2022).

Although most traditional Spanish grammars assume that both forms are rarely interchangeable, these sources have not inquired how these past tenses are represented in corpus data. Regarding the use of the Preterite and the Imperfect, traditional Spanish grammars refer us to its use within different linguistic factors as well as what the message aims to convey. In addition, there are some discrepancies regarding the theories of its use with the actual use of these past forms. For example, Delgado-Díaz (2014) compared the traditional grammars descriptions of the use of the Preterite and Imperfect versus the actual use by Puerto Rican speakers. The author found inconsistencies between some inconsistencies between its prescribed use and its actual use in the Puerto Rican variety.

² The examples gathered from the CESA corpus have a parenthesis under the examples that start with (CESA)

³ The examples gathered from the CEPA corpus have a parenthesis under the examples that start with (CEPA) or (PhoeCo)

Moreover, Delgado-Díaz digs further into the use of the Preterite and Imperfect in a different study and showed that dialectal differences exist with regard to the prediction of the Preterite and Imperfect in Argentinean and Puerto Rican Spanish (Delgado-Díaz, 2018).

Taking this evidence of variation combined with the findings of varieties of Spanish in the U.S spoken in Los Angeles (Silva-Corvalán, 1994, 2003, 2014) and New York (Zentella, 1997), where the data have shown that the Preterite and Imperfect opposition is not as clear as the Spanish grammars have illustrated, variation may not only exist in different monolingual varieties but it must also exist in other varieties of Spanish in the U.S, specifically in areas with a high degree of language contact. Furthermore, in an effort to deviate from the preconceived decisions the grammar manuals present, this dissertation analyzes the Preterite and Imperfect past forms to find their obligatory uses and contexts in the Spanish spoken by three generations of bilinguals in Arizona.

The Spanish past tense

The past tense has been defined as the “location in time prior to the present moment” (Comrie, 1985, pp. 41) and as such, there are different forms that can be used in order to refer to the past. In Spanish, there are five forms of the past in the indicative mood. The five are -- *el pretérito imperfecto, el pretérito indefinido, el pretérito perfecto, el pretérito pluscuamperfecto y el pretérito anterior* (Alcina Franch & Blecua, 1975; Bello, 1847; Gili Gaya, 1961). The most common forms to express the past are the

indefinite Preterite, also simply known as the Preterite (1) and the Imperfect (2) (Comrie, 1976; Bybee, Perkins & Pagliuca, 1994).

(1) El año pasado cuando..., sí el año pasado, cuando **hice** mis taxes, no **agarré** mucho dinero porque apenas **empecé** con el trabajo que tengo ahorita.

(CESA049_619)

(2) **Jugábamos** fijate **jugábamos** football en la pura calle, o sea football y **era** tackle football.

(CESA036_657)

The main difference between the Preterite and the Imperfect is its aspectual features (Comrie, 1976). Spanish marks whether the past state or event is of perfective aspect or imperfective aspect and this it is morphologically expressed on the verb. This type of characteristic, while not morphologically marked in English, however, English has ways of expressing this feature when necessary.

Aspectual domains of the Spanish past tense

In Comrie's (1976) seminal book, the author provided a graph of proposed aspectual distributions. As Figure 1 shows, the classifications are first divided between perfective and imperfective. Comrie (1976) states that "Perfectivity indicates the view of a situation as a single whole, without distinction of the various separate phases that make up a situation; while the imperfective pays essential attention to the internal structure of

the situation” (pp. 16). The figure shows that while the perfective aspect ends there, the imperfective aspect covers different functions, habitual and continuous. The imperfective feature encompasses a “reference to the internal temporal structure of a situation, viewing a situation from within” (pp. 24). The *habitual imperfective* function refers to an event that occurred more than once and has been repeated during a long period of time. The continuous function has *progressive* and *non-progressive* features. While the *progressive* function refers to a continuous event in the reference time (Bybee et al., 1994), the *non-progressive* hints at a nearly permanent state of affairs (Comrie, 1976). The *progressive* function has typically been observed with adverbs such as *mientras*, as well as with the *estar* + GERUND periphrasis construction which emphasized the ongoing event.

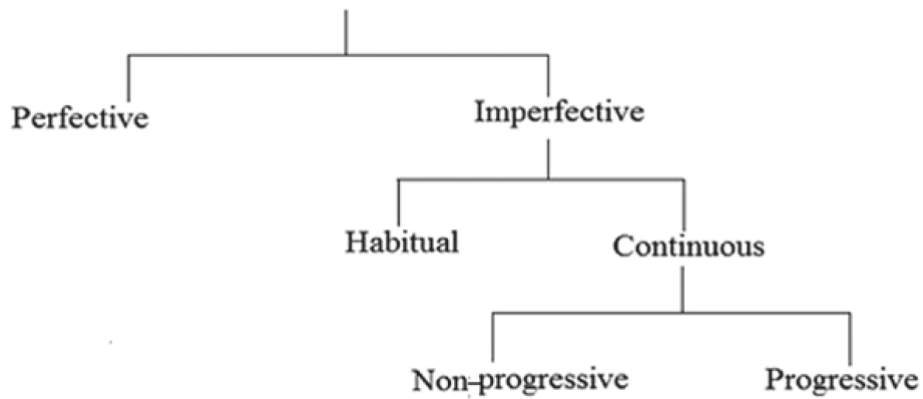


Figure 1. Aspectual distributions (Comrie, 1976:25)

While many of the aspectual distribution categories appear unambiguous, Bybee et al. (1994) argued that the categories continuous and non-progressive were impossible

to distinguish in their data. Therefore, the authors rejected both categories and continued to only discuss the perfective, habitual and progressive aspectual features.

Perhaps the most discussed past tense grammar points in Spanish are the Preterite and the Imperfect. Many textbooks focus on making a clear morphological distinction between the two as early as the first level of Spanish as a second language as well as the first level of Spanish as a heritage language. These distinctions typically focus on how and when to utilize one form over the other. As previously discussed, the imperfective aspect is not as straightforward as the perfective aspect as there are many subcategories that can be included as part of an imperfective aspect. However, most textbooks typically assign a durative aspect to the Imperfect and punctual aspect to the Preterite. In addition, the use of the Preterite typically signals that there is a specific time when the state or action has ended, much like the definition provided earlier for the perfective aspect. Meanwhile, the Imperfect can indicate a habitual or progressive action. An example of use of the Preterite is shown in (3), where the morphology of the verb indicates perfectivity which signals the action of buying the ticket has been fully completed. On the other hand, paradigm (4) exemplifies the Imperfect past form with a habitual imperfective function. This example suggests that this participant would not copy nor cheat off his classmates throughout his schooling in the past. More specifically, the speaker is denying habitual actions he did not take part of.

- (3) Bueno, pues **compré** tickets para ir a ver The Weekend en abril, pero fue cancelado.

(CEPAHM_L133)

(4) La verdad, nunca, nunca **copiaba** ni **cheateaba** para hacer cosas así.

(CEPAHM_L113)

Example (4) is particularly interesting due to the fact that the verb ‘to cheat’ was borrowed from English and conjugated according to Spanish morphological paradigms in order to interpret the Imperfect past. As one can imagine, this type of phenomenon can only naturally occur in a Spanish English bilingual setting. Paradigms such as (4) add to the necessity of studying the Preterite and Imperfect in real speech, in this case through sociolinguistic interviews, a form such as *cheateaba* would most likely be excluded from analyses that take solely experimental approaches. Examples like this are important as they highlight the reality of being bilingual in the United States. Given the multilingual and multicultural context of the United States, specifically the Southwest, the contact that Spanish and English share has afforded the emergence of many linguistic phenomena that although could exist, may not be evident in the linguistic systems of monolingual varieties of either language. Given that Spanish English bilinguals live in a constant state of language contact, it is also relevant to summarize what the past looks like in English.

Aspectual domains of the English past tense

Unlike Spanish, English is characterized by one simple past. Perfectivity versus imperfectivity are not distinguished morphologically and these specific opposition has not been grammaticalized (Comrie, 1976). The simple past in English assumes a perfective aspect, which indicates that an exact way to encode the Imperfect aspect of the

simple past is not morphologically provided by the language. Instead, auxiliary verbs such as ‘used to’, ‘would’ or can be used to indicate imperfectivity and habitualness in the English language (Montrul & Slabakova 2002) such as example (4b). 4b is the translation of the previous example 4a showed in section 1.1. This translation was provided to observe the differences between the aspectual domains of Spanish and English. 4b shows that in English, in order to show a more habitual aspect to an event, an auxiliary verb is needed.

(4a) La verdad, nunca, nunca **copiaba** ni **cheateaba** para hacer cosas así.

(HM_L113)

(4b) Honestly, I never (**used to copy**)/**copied** nor (**used to cheat**)/**cheated** to do things like that

While the perfectivity and imperfectivity opposition has not been grammaticalized in English, the progressive versus non-progressive opposition with non-stative verbs and when excluding the habitual meaning is a comparable opposition in English to the perfectivity versus imperfectivity opposition in Spanish (Comrie, 1976). (5) and (6) show examples of how the non-progressive and progressive opposition would be represented in the English past tense, respectively.

(5) Juan **ate** his food

(6) Juan **was eating** his food

Understanding how bilingual speakers can interpret the past is an important step to analyzing both the Preterite and the Imperfect. This understanding comes from the concept of translanguaging which can be defined as a “complex language practice of plurilingual individuals and communities” (García & Wei, 2014, pp. 20) or more specifically “the ability of multilingual speakers to shuttle between languages, treating the diverse languages that form their repertoire as an integrated system” (Canagarajah, 2011, pp. 401) As translanguaging takes place with such bilinguals, it can be assumed that all of these forms play a role in bilingual speakers when these speakers are expressing the past tense. Through translanguaging and straying away from preconceived notions that may be present with the Preterite and the Imperfect the data can show what the relationships between form and function exist between these two verb forms (Preterite and Imperfect) and its semantic categories.

Languages in Contact

Over the years many definitions of language contact have been proposed. These definitions for the most part express the same idea. One of the pioneering and seminal works on the topic defines languages to be in contact “if they are used alternately by the same persons” (Weinreich, 1979, pp. 1). This definition assigns the bilingual speaker as the main focus of language contact. However, as Thomason (2010) mentions, there should be an emphasis on the speakers’ use and interactions of both languages since as sociolinguists our interest is typically society and not necessarily the individual as the

focus. My perspective on the results of language contact, especially with matters that encompass language change, aligns with that of Hein and Kuteva (2005) who explain that while speakers in situations of language contact may create “new structures drawing on universal strategies of conceptualization” these characteristics are not viewed as an Imperfect use of a language (Hein & Kuteva, 2005, pp. 36). Instead, they are simply the manifestation of the realities of a bilingual speaker in a contact situation.

To consider that language contact can lead to language variation and change does not necessarily mean that structures of one language are interfering or being transferred to the other or vice versa. While language contact can sometimes ease the acceleration of processes already underway in monolingual varieties, it is also pertinent to consider other possibilities to contact when referring to language variation and change (Poplack & Levery, 2010). In case a variety spoken by bilinguals present a change in progress, it is important to assess whether the changes were driven by internal versus external tendencies (Carvalho, 2016). Internal tendencies refer to the language’s own structure changes while external tendencies indicate changes caused by another language. Like many other sociolinguistic studies, the goal of this research is to present a synchronic analysis of the Preterite and Imperfect in the Spanish English bilingual community in Arizona. The next subsection presents background information on the state of Arizona and why it is considered an area with intensive language contact.

Spanish speakers in Arizona

While the anti-immigrant and anti-bilingual ideologies present in the state cannot be denied (Cashman, 2006, 2009), Spanish speakers in Arizona are a stable and growing

community. The presence of Spanish and Hispanic/Latine cultures in the Southwest of the US dates back to the 16th century when the Spanish expeditions arrived in the area (Silva-Corvalán, 2004). The Southwest is being defined following Silva-Corvalán's grouping which includes Arizona, California, Colorado, New Mexico and Texas (Silva-Corvalán, 2004). More recently, some of the largest concentrations of Hispanics and Latines are found in the Southwest. Seven out of the ten U.S metropolitan areas with the most Hispanic population are part of the Southwest; Phoenix, Arizona being number 8th on the list with 1,464,000 Hispanic residents (Pew Hispanic Research Center, 2017).

In Arizona, there is a population estimate of 7,278,717 inhabitants. Out of the total population, 31.7% identified as Hispanic or Latine (United States Census Bureau, 2019). Furthermore, 27.8% indicated they were of Mexican descent. Arizona forms part of the Southwest of the US which has some of the largest population of Spanish speakers. In fact, the number of Spanish speakers in Arizona (27.2%) is higher than the national percentage of Spanish speakers (13.5%) (United States Census Bureau, 2019).

The state of Arizona borders Mexico, which allows the population to go back and forth and increases the interactions these speakers have with friends or relatives (Cerrón-Palomino, 2016). Due to this proximity, it could be suggested that this particular community receives a greater amount of input of their own variety or a similar one. While the proximity to their heritage varieties is relevant, it is also pertinent to conduct studies that allow researchers to observe US Spanish speakers varieties without the necessity to compare to its monolingual counterparts.

The corpora utilized in the present variationist analysis only includes bilingual speakers living in the state of Arizona. Data collected by the American Community Survey shows that 27.2% of the population in Arizona speak a second language. Out of this 27.2%, 20.3% of these are Spanish English bilingual speakers (United States Census Bureau, 2019). This equals to approximately 1,477,579 of Spanish English bilinguals in the state of Arizona. Despite the numerous bilingual and multilingual speakers in the US, and more specifically Arizona, bilingualism has been discouraged in many places for a very long time. In the US, many bilingual speakers still suffer from discrimination and confrontation due to anti bilingualism ideologies. This is evident in the history of state legislation like Arizona's 2010 State Bill 1070, which instituted an anti-immigration policy, and Proposition 203, which eliminated "bilingual education in state-funded schools" (Cashman, 2006, pp. 42). Proposition 106 made English the official language of business in Arizona. This reality is important to highlight in order to take into consideration the struggles and strength of participants in our studies.

The Spanish English bilinguals that form part of the data analyzed in this dissertation include a specific but heterogeneous group of speakers. However, they all form part of a language learning experience described by the concept of differential bilingualism (Aparicio, 1998). This phenomenon refers to the unequal value that is given to bilingual speakers by the dominant English monolingual speakers in society. On one hand, the Anglo bilingual speakers who are privileged to experience elite bilingualism, which encompasses learning their second language as an extra course or through private means, are typically viewed as people whose bilingualism is resourceful and an asset to society. In contrast, bilingual speakers who speak the minoritized language first or grew

up with both languages being spoken at home, and most likely come from immigrant backgrounds, tend to be perceived as people whose bilingualism is a disadvantage. All of the speakers included in this dissertation could form part of the latter description and be seen as such to society.

Due to the dominance of English in the US, those who use English as a second language and those who grew up hearing a minority or immigrant language are marginalized and led to believe that monolingualism is the norm (Leeman, 2012), which can lead to language loss. Additionally, according to Fairclough (2016), those bilinguals who learn their language at home frequently acquire the stigmatized and ideologically laden home variety of the language (Beaudrie et al., 2014; Potowski, 2002; Sánchez-Muñoz, 2016).

The dominant English speakers in society typically approach languages from several perspectives, one of which is language-as-problem (Ruíz, 1984). The non-dominant language speakers such as the immigrant Spanish speakers in the US face linguistic prejudice because English speakers frequently consider minoritized languages as a problem (Cashman, 2006), regardless of the variety spoken or whether a particular dialect is spoken by the speaker of that language (Lippi-Green, 2004).

In addition, understanding that bilingual speakers are not and should not be considered to be two monolinguals in one is crucial (Grosjean, 1989). Within the bilingual range of the participants that form part of this study, there are bilinguals that acquired their second language, in most of the cases English, at different stages of their life. Some of the speakers grew up speaking Spanish most of their life and acquired

English in adulthood, while some speakers grew up speaking both languages. A further description of the heritage speakers is provided below.

Heritage Speakers of Spanish and Heritage Language Learners

As mentioned above the bilingual corpora included a range of bilinguals. Within this range are included heritage language speakers. Valdés (2001) defines a Heritage Language (HL) in the United States as “all non-English languages including those spoken by native American peoples” (pp. 39). Therefore, a heritage speaker of Spanish in the US may be a speaker who grew up in the US surrounded by Spanish.

Heritage speakers (HS) are highlighted in this dissertation due to the pedagogical implications that this study aims to highlight in the discussion section (Chapter 4). When a HS has the opportunity and makes the decision to enroll in HL course, this speaker is considered an HL learner. The most prevalent definition in the field of HL studies in the United States defines a HL learner as an individual who “is raised in a home where a non-English language is spoken. The student may speak or merely understand the heritage language and be, to some degree, bilingual in English and the heritage language” (Valdés, 2000, pp.1). In addition to being considered the most cited, also turns out to be a narrow definition since it assumes that these students have developed some competence in the heritage language (Polinsky & Kagan, 2007). Nevertheless, there are heritage language programs that include students who have strong language connections but have not been exposed to the heritage language at home. The different experiences and backgrounds of the HL learners bring a heterogeneity to the classroom with respect to their previous knowledge, as well as turn out to be a challenge for the instructor who, as

will be seen later, must develop a series of dynamics and strategies to make the various topics addressed in the HL classroom effective.

In fact, Beaudrie, Ducar and Potowski (2014) mention that HL learners are a heterogeneous group that differ from second language learners (L2) according to the five dimensions of student diversity: historical, linguistic, educational, affective, and cultural. These dimensions play a central role not only in the linguistic competence of the student, but also in their linguistic motivation. For these reasons, HL courses have become commonplace in institutions around the country with the purpose of developing the strengths of students and assisting them in acquiring continuation of their HL (Beaudrie, 2012; Beaudrie et al., 2014). In fact, in 2012, 40% of the universities in the United States already offered Spanish as a Heritage Language courses (SHL). This was a 45% increase since 1990 (Beaudrie, 2012). Likewise, these programs have encouraged researchers in the field to analyze the role of grammar in these courses and the benefit this would bring to the SHL classroom.

Rationale

The goal of this dissertation is to analyze the use of the Preterite and the Imperfect in three generations of Spanish English bilingual speakers in Arizona by using oral corpus data. This goal aims to cover two main gaps in the literature. First, most of the previous research that has studied these two past forms approach the variable from mainly experimental methodologies that aim to demonstrate how certain speakers choose the “correct” forms for both the Preterite and the Imperfect (e.g., Cuza, et al., 2013; Cuza & Miller, 2015; Montrul, 2002, 2009) (see Chapter 2). While experimental data may yield

results efficiently for the purposes of language acquisition, experimental data is not very representative of natural speech and have its limits (Poplack & Dion, 2009). Therefore, given the heterogeneity found among HSs, a sociolinguistic approach that analyzes what linguistic and extra-linguistic factors favor and disfavor the use of both past tenses is necessary.

Second, more recent research (e.g., Delgado-Díaz, 2014, 2018) has shown that the dichotomy of the Preterite and the Imperfect is not as clear as traditional Spanish grammars had highlighted in the past, and how it is still to this day highlighted in many of the language classrooms. There is still a lack of empirical research on these past forms using corpus data. Therefore, empirical research that is able to demonstrate what favors the use of one past form over the other based on corpus data is also necessary.

While a few studies such as Silva-Corvalán (1994, 2003) and Zentella (1997) have highlighted the aspectual overlaps regarding the Preterite and the Imperfect in their qualitative work, the phenomena have not been studied through quantitative methods. To date, there are no corpus studies that quantitatively and qualitatively provide complete focus on these two past tenses. Moreover, there are no studies that focus on these variables in the state of Arizona. A further goal of this study is to understand these two past forms in the speech of Spanish and English bilinguals in Arizona in order to gain an understanding if this use can be considered a characteristic of this variety of Spanish in the US. More importantly, the field of Spanish in the U.S will benefit from a study that uses a variationist lens to demonstrate what factors favor the use of this variable in the speech of bilingual speakers.

Lastly, one of the main goals of this dissertation is to gain detailed insight into the synchronic workings of the Spanish spoken in Arizona by comparing the use of the Preterite and Imperfect between three generations. This comparison is important as the literature has shown that the further the generation the more language loss that occurs. In addition, this comparison can inform us if the factors that favor the Preterite and the Imperfect carry similarities through generations as in if it is presenting a possibly characteristic of US Spanish in Arizona or if a change might be in progress.

Organization of the Dissertation

This section provides a brief outline of the dissertation to help the reader. The present dissertation has five chapters. Chapter 1 introduces the reader to the Preterite and Imperfect in the Spanish language, its uses, as well as misconceptions. In addition, this chapter details what aspect is and how it is being defined. Lastly, it presents the reader with a brief description of how this dissertation is organized. Chapter 2 takes the reader through a review of the literature pertinent to the topic of the Preterite and Imperfect use in Spanish and focuses on presenting previous works that have focused on monolingual Spanish speakers as well as bilingual English Spanish speakers.

Chapter 3 focuses on describing the methodology utilized in the present study. This chapter describes the participants and data utilized, the envelope of variation, the linguistic and extra-linguistic factors, and the data analysis. This chapter provides a variety of examples to accompany every section. Chapter 4 presents the quantitative results and is corroborated by qualitative data in form of excerpts from the participants' interviews. Chapter 4 also provides the discussion section which includes implication of

these results for future research and for the Spanish language classroom, especially the Spanish as heritage language classroom. Lastly, Chapter 5 provides a conclusion, the contribution this dissertation makes, the limitations, and suggestions for future research.

CHAPTER 2

LITERATURE REVIEW

This chapter presents an overview of previous studies that have examined the past tense in Spanish, especially those that have examined the uses of the Preterite and Imperfect. The first part provides a brief overview of the sociolinguistic research trajectory that has led to morphosyntactic variables such as the Preterite and the Imperfect. This is followed by a discussion on the rise and importance of variation studies. Next, I will present what previous traditional Spanish grammars have discussed regarding the Preterite and the Imperfect. Fourth, this chapter reviews previous research on the use of these past forms by monolingual Spanish speakers. Finally, this chapter presents a review of the literature concerning the use of the past tense by Spanish English bilinguals. While most of these previous studies take a Second Language Acquisition (SLA) perspective and for the most part use experimental methodologies, a special emphasis is given to the very few studies that have observed these variables utilizing corpus data and sociolinguistic interviews.

The previous work conducted on the Imperfect and Preterite forms is extensive as they are perhaps the most studied Spanish past forms. However, these past forms have not quite been explored as a variable in past studies. In addition, not much empirical research has been conducted regarding the actual use of these two past tenses in both monolingual and bilingual communities that moves away from preconceived notions of how each past form is supposed to behave. In addition, most of the work that has used the Preterite and Imperfect as variables are Second Language Acquisition studies. In order to create a better understanding of what previous studies have shown, I will summarize the

relevant literature, starting with the traditional grammars' account of the use of the Preterite and the Imperfect. This will be followed by exemplifying the research conducted within the sphere of corpus linguistics regarding monolingual varieties of Spanish. In this section, I will outline the studies that have analyzed this variable within Spanish English bilingual communities. Lastly, I present pertinent concepts that are involved when studying participants from different generations such as language maintenance, language shift and language loss.

Morphosyntactic variables

Most of the pioneering works in the field of sociolinguistics focused on the study of phonological variables. Following the definition of a linguistic variable, it was easy to propose that two phonological variants had the same referential meaning. A great example is the seminal work of Labov (1963, 1973) where he observed an in-progress sound change on Martha's Vineyard: the (aw) and (ay). These two diphthongs were being centered, creating two diphthong alternatives in the speech of the native population. Labov found that the centralization of the diphthong was the most obvious in the speakers that belonged to the 31-45 age group. He explained that this centralization was a way the vineyarders expressed their belonging to the island, a sentiment these speakers also shared. Variation between two or more phonological variants became a trending topic in the studies of language variation and sociolinguistics.

Shortly after, Sankoff (1973) proposed an extension of the original sociolinguistic framework that expanded the scope of the sociolinguistic variable. This extension would encompass the study of syntactic and semantic variables. This meant extending the

quantitative work on language variation to variables that examine lexical, syntactic, and morphological variation. She presents studies that analyze such types of variation in order to demonstrate that variability was present beyond phonological variables. The sample variables presented were (1) the future marker in New Guinea Tok Pisin, (2) the complementizer *que* in Montreal French, and (3) the Montreal French indefinite *on*.

Sankoff found some syntactic constraints when using the future marker in New Guinea Tok Pisin. For example, all the pronouns except for the third person singular pronoun follow the future marker *bai*. Additionally, she found that the future marker *bai* functioned to mark specific emphasis and exclusivity. With regards to the complementizer *que* in Montreal French, she observed that the presence or omission of *que* depended on the grammatical construction of Montreal French. She shows that speakers had similar levels of acceptability to different *que* insertions. For example, while *quand c'est que* 'when is that' was acceptable, **comme c'est que* 'as that is' was not. Lastly, she presented part of Laberge's (1977) work on the Montreal French indefinite *on*. Findings showed that the pronoun *on* was preferred by both male and female speakers over the age of 40. Moreover, within the younger speaker group, a big sex difference was present: younger women preferred *on* to a greater extent than men. The percentages also showed that younger women used the pronoun *on* more than the males and females in the older group. By providing examples on non-phonological variation, Sankoff highlighted how relevant analyzing morpho-syntactic variables can be in order to understand the grammar of a particular variety, as well as language use across factors such as sex and age. She concluded by stating that "the extension of probabilistic considerations from phonology to syntax is not a conceptually difficult jump" (Sankoff, 1973, pp.92).

Although it faced some skepticism from some researchers (e.g., Lavandera, 1982), the expansion of the sociolinguistic variable was supported by many, as it expanded the scope of the sociolinguistic variable. The rise of variationist studies on morphosyntactic features opened doors to studies such as the study presented in this present dissertation as this new focus was also spread to the study of linguistic variables in other languages such as Spanish. Variationist sociolinguistics with a concentration on the study of Spanish phenomena rose in the second half of the 20th century (Díaz-Campos, 2022). These variationist studies have continued to be useful to demonstrate not only how language is spoken in different Spanish speech communities but also to aid the understanding on how language is evolving.

The Preterite and Imperfect in monolingual Spanish speakers

Traditional Spanish grammars make it seem very easy to distinguish between the use of the Preterite and the Imperfect in monolingual speech (e.g., Alcina Franch & Bleca, 1979). Consequently, it is assumed that their use is not interchangeable and the use of the Imperfect/Preterite would not be strictly considered a morphological variable of study in a monolingual or bilingual setting. However, there are some authors that have gone beyond the Spanish grammars' account and have analyzed oral data in order to get a better understanding of its actual usage in different varieties of Spanish.

Silva-Corvalán (1983) studied the narratives of 17 Chilean monolingual speakers of Spanish and three Mexican monolingual speakers of Spanish. The author analyzed tense and aspect in oral narratives with the goal of quantitatively and qualitatively observing the use of the historical present, Preterite, and Imperfect. The results illustrated

that the type of information produced was an important factor when using these different forms. Speakers preferred to use the Imperfect with background information while the Preterite was preferred to be used with foreground information (Silva-Corvalán, 1983).

Other studies such as Schwenter and Torres Cacoullos (2008) and Howe and Schwenter (2008) compared the Preterite and the Present Perfect in monolingual varieties. Schwenter and Torres Cacoullos compared the use of the Preterite and Present Perfect in Mexican and Peninsular varieties of Spanish. They found that in the Peninsular variety the use of the Present Perfect is becoming the default past tense use, which was not the case in Mexican varieties. Regarding the Preterite, it appeared that in the peninsular variety the Preterite was mainly used when the temporal context was the furthest in the past. Additionally, while the preferred use to refer to a perfective past is the Preterite in Mexico, it is more popular to use the Present Perfect in peninsular Spanish. Similarly, Howe and Schwenter (2008) analyzed the speech of Lima's 1980s monolingual norm (Caravedo, 1989) and compared the results of the Preterite and Present Perfect with a variety from Madrid and a variety from Mexico. The study's main results were the comparison of use between the Lima and Mexico varieties regarding the Present Perfect, which has been noted to be preferred in the Madrid variety with the highest rate. While this study focused mainly on the use of the Present Perfect, the results revealed that the use of the Present Perfect in Lima speakers did not follow a similar grammaticalization trend as the Peninsular variety (Howe & Schwenter, 2008). In addition, the data shows that the Preterite is highly favored (73.6%) in the Spanish of Lima. This use is compared to Lope Blanch's (1976) study in Mexico City (85.2%). While the main focus of both studies was the Present Perfect, both studies demonstrated

that between Latin American and peninsular varieties, the former prefer the Preterite to refer to the past whereas the latter varieties favor the Present Perfect.

More recently, Delgado-Díaz (2014) analyzed data from monolingual Puerto Rican speakers from the PRESEEA (Proyecto para el Estudio Sociolingüístico del Español de España y de América) corpus. The author studied the speech of 12 Puerto Ricans according to generation: first, second and third. Findings demonstrate that the analysis of natural speech does not fully coincide with what studies on the experimental data have shown. The factors he found to be significant include lexical semantics, the specificity of the event, the time frame of reference, and age (Delgado-Díaz, 2014). More specifically, the author found that the verbs of accomplishment favor the use of the Imperfect which admits some degree of variation due to Spanish grammars typically assigning a telic duration to favor the use of the Preterite. Delgado-Díaz (2014) highlighted that only lexical semantics and specificity of the event have been shown as significant with regard to the Imperfect but not much has been found on age and time frame of reference except for his study.

More recently, Delgado-Díaz (2018) investigated which linguistic factors predicted the use of the Preterite and the Imperfect in two distinct varieties of Spanish, Puerto Rican, and Argentinian. The goal was to determine if dialectal variation existed within the use of both past tenses. A total of 12 participants' data were analyzed: six participants from Puerto Rico and six from Buenos Aires. A total of 411 tokens were produced by the Puerto Rican speakers versus 295 from Buenos Aires. Interestingly, more instances of the Imperfect were produced within the Puerto Rican data while the Buenos Aires data produced more instances of the Preterite. Regarding linguistic factors,

the results suggest that more variation might be present than what the traditional Spanish grammars may have initially thought (Delgado-Díaz, 2018). More specifically, specificity of the event, the temporal frame of reference, and the lexical semantics were the factors that determined the use of the Preterite and Imperfect in Puerto Rican Spanish while the specificity of the event, the lexical semantics, the type of information, and the plurality of the object were the relevant factors in Buenos Aires Spanish.

Delgado-Díaz (2021) continues his work on the Spanish past tense by recently publishing a book analyzing the diachronic change of the past tense across the Spanish language through literary works from the following periods of time: Medieval Spanish, Golden Age Spanish, and Modern Spanish (Following Torres Cacoullos, 2012). His motivation to carry out this study stems on previous studies that have shown aspectual overlaps (i.e., an ongoing event in the past) (Bybee, Perkins, & Pagliuca, 1994; Delgado-Díaz, in press; Lamanna, 2008, 2012). His results demonstrated that the different past forms were not as rigid as they originally were thought to be. The majority of his data were composed of the Preterite and the Imperfect which supports the statement that the Preterite and the Imperfect are the most popular forms of the past tense.

Diachronically, the data suggested that the shifts from one form to another could have been due to the appearance of the past progressive forms, as many forms are competing with these progressive structures. Perhaps more specific to the current study, the author concluded that the Preterite and the Imperfect are not opposites (Delgado-Díaz, 2021, pp. 126). He states that both the Preterite and the Imperfect each carry semantic functions that the other does not. Therefore, they cannot be considered exact opposites. Regarding the frequency of these two specific past forms, the results

demonstrated that the Imperfect form had increased over time. The Imperfect represented 21.9% of the data in Medieval Spanish, 33.4% in Golden Age Spanish, and 44.7% in Modern Spanish. However, the opposite pattern occurred with the Preterite, as the Preterite formed 34.4% of the data in Medieval Spanish, 33% in Golden Age Spanish, and 32.6% in Modern Spanish. The decrease of the Preterite was not as strong as the increase of the Imperfect. Taking these results alone, it appears that the Imperfect has gained more use as time goes by which could have moved into canonical territory of the Preterite. The factors that favored the use of the Imperfect in Modern Spanish were aspectual function, type of information, frame of temporal reference and priming, while the factors that favored the use of the Preterite were aspectual function and type of information.

Although not much research has been conducted on the variation of the past tense in monolingual speech, the work highlighted on monolingual speakers demonstrates that 1) the aspectual characteristics previously established by traditional Spanish grammars are not as rigid as they may have seemed and 2) beyond what has been stated by conventional grammars, there may be additional factors that influence whether one form of the past is preferred over another. As previously stated, most of the work investigating the use of the Preterite and the Imperfect have been conducted from an SLA perspective where experimental methodologies were employed. Experimental data have limitations when it comes to analyzing what factors may favor the use of these types of variables. In the section below, I will provide a review of studies that have investigated this variable in Spanish English bilingual speech.

The Preterite and Imperfect in bilingual speakers in the US

Most of the work observing the Preterite/Imperfect variables in Spanish Heritage Speakers comes from a generativist perspective with a focus on the level of acquisition and competence of the past tense. One of the most prominent studies is Montrul (2002). The author analyzes whether the age of onset of bilingualism has an effect on the accuracy of Preterite and Imperfect form use in narratives as well as the interpretation of the Imperfect and preterit forms by adult Spanish Heritage Speakers. Montrul compared a monolingual group with three bilingual groups (simultaneous, early child L2, late child L2). With regards to accuracy, a written morphology recognition task showed “non-native-like” performance is attested with individuals in the three groups, many more divergent cases are found in the two earliest groups (simultaneous bilinguals and early child L2 learners)” (Montrul, 2002, pp. 49).

Meanwhile, an oral production task revealed that a preference for the Imperfect form with atelic verbs (stative, activity) and a preference for the Preterite with telic verbs (accomplishments and achievements). Additionally, not much of a difference between the monolingual and bilingual groups was observed. With regards to the interpretation of the Imperfect and Preterite, a sentence conjunction task revealed that all bilingual groups behaved differently than the monolingual group with regards to the stative verbs in both the Preterite and Imperfect. However, a truth value judgment task shows that only the late L2 child learners performed like the monolingual speakers. A summary of her results leads the author to suggest “incomplete acquisition in the case of simultaneous bilinguals

and early child L2 learners and attrition in the case of late child L2 learners affect more profoundly stative verbs in the Preterite.” (Montrul, 2002, pp. 57)

In another study, Montrul (2009) analyzes tense, aspect, and mood of early bilingual adults. In the tasks that specifically focus on Preterite and Imperfect she compares 23 “fluent native speakers” with 65 heritage speakers of Spanish. All participants completed written and oral tasks, all tasks were experimental or semi-experimental. The oral retelling task showed that all speakers produced most instances of the Preterite. Additionally, “native speakers, intermediate, and advanced heritage speakers did not differ from each other, but they all produced more Imperfect forms than the low proficiency heritage speakers” (pp. 251). The author also highlights the “errors” the participants produced. These “errors” show instances where the Preterite was used instead of the Imperfect or vice versa. For example, in example (7) the participant with intermediate proficiency in Spanish uses *decía adiós* which in this study is considered an error. However, from a more descriptive perspective, this would be considered a perfective aspectual feature while using the Imperfect. According to this study, there was an overall higher rate for Imperfect “errors” than Preterite “errors.” The written task findings reveal that overall, there was a “lower accuracy” of the Imperfect.

(7) “Al final, la abuelita, Caperucita Roja y el señor vieron caminar al lobo. La abuelita le ***decía** adiós a Caperucita y al señor. El fin.”

‘In the end, the granny, Little Red Riding Hood and the man saw the wolf walk.

The grandmother said goodbye to Little Red Riding Hood and the man. The end.’

(Montrul, 2009, pp. 250)

Cuza, et al., (2013) conducted a cross-sectional study that examined the productions of the past tense in 13 simultaneous Spanish English bilinguals, a group that was divided into younger and older bilingual learners, and 11 adult Heritage Speakers (HSs) of Spanish. The authors compare data gathered from a story-telling task with data from monolingual children and adults of the same age as the HSs. The results demonstrated that both the bilingual and monolingual children used the Preterite at a higher rate than the Imperfect. However, the differences for past production were significant between the monolingual and bilingual speakers. Results showed that the monolingual adults' use of the Preterite and Imperfect is somewhat more balanced than that of bilingual adults. In addition, they found that the older bilingual children used a higher proportion of Preterite forms than the younger group. However, the use of the Imperfect remained similar between both younger and older groups. The authors point out that this "overuse" of the Preterite differs from its use in adult bilingual speakers. They argue that "the overuse of the Preterite among older children stems from the transfer of the aspectual selectional properties that tense heads are able to select in English" (Cuza, et al., 2013, pp. 212). They hypothesize that since the English Preterite is neutral, the speakers almost mimic this and choose the Preterite as that neutral form. They point out that the Imperfect is more semantically marked or complex therefore it is expected to be lost first (Cuza, et al., 2013; Silva-Corvalán, 2014). The authors describe the use of the HS Imperfect as something that "remains incompletely acquired" (pp. 216) solely based on the fact that the HSs both children and adults keep using the Imperfect, while the monolinguals prefer to use the present and Preterite forms when producing a

narrative. Additionally, they did not find any instances of “overgeneralization” with either the Preterite or Imperfect. While they do point out that these results should be taken cautiously, due to the fact that the same methodology was not used in both participant groups. However, this study provides insights into what participants, children, adults, monolinguals, and bilinguals produce when it comes to the Preterite and the Imperfect.

In a follow-up study, Cuza and Miller (2015) analyze the speech of 19 Spanish English bilingual children and 12 of the children’s parents. However, in this study, the authors analyzed if language dominance and target performance were relevant factors. The methodology was also different, in that instead of a story re-telling task, the authors used a question-after-story task. Results showed a very low proportion of the Imperfect form. Furthermore, both children and adults demonstrated a high level of production of the Preterite, except with stative predicates. However, the authors mention that the ambiguity of intended interpretation may have had an effect. Results also show that target performance did not correlate with language dominance. However, although not significant, the older children outperformed the younger children despite the longer contact with English the older children may have had.

Another study that focused on HS children is Silva-Corvalán (2003). She studied the speech of six bilingual children ages 5-11. In her analysis, she compared the production rates of tense and aspect to 10 second and third generation participants from her earlier work in Los Angeles (Silva-Corvalán, 1994). In her 1994 study, which will later be discussed in more detail, participants from the second and third generation used Imperfective morphology in perfective contexts with stative verbs. Through recordings of

the children, Silva-Corvalán noted that three of the children used the Imperfect in all contexts where the meaning was Imperfective. However, the other three children did not always use the Imperfect in Imperfective contexts. Furthermore, the analysis showed that only one child used the Preterite in all the perfective contexts, while the rest of the children favored the Imperfect even in some perfective contexts. The author suggests that this “simplification” results from “an interrupted acquisition process resulting from reduced Spanish input” (Silva-Corvalán, 2003, pp. 393). On the other hand, she also says that by the age of the children in the lower age spectrum, children typically do not have a “complete” sense of tense, mood and aspect in Spanish.

As previously described, the studies discussed until now have either approached their analysis with experimental tasks and with the purpose of rating participant/speaker production and comparing it with monolingual varieties; or have focused mainly on children’s production. The following two studies come from two seminal books that observed the Preterite/Imperfect opposition from a sociolinguistic perspective and from corpus data: Silva-Corvalán (1994) and Zentella (1997).

In her book on language contact and change, Silva-Corvalán (1994) examines the hypothesis that Mexican American bilinguals in a high multi-ethnic area such as Los Angeles develop strategies for how to deal with the cognitive load of their linguistic systems due to the language contact situation. Los Angeles, at the time of the data gathering, as well as now, is a city with a large Hispanic and Latinx population. Particularly, at the time of the study, the eastside of Los Angeles had a population where at least 40% of the residents claimed to speak Spanish as their home language. Therefore, extensive contact between Spanish and English exists. Silva-Corvalán conducted 50

audio-recorded conversations with Mexican Americans in the eastern area of Los Angeles. Her participants were classified into three different groups, Group 1, Group 2, and Group 3.

Participants were categorized intergenerationally and based on how long they had lived in the US Group 1 included speakers who were born in Mexico who had moved to the U.S after age 11; Group 2 included participants who were either born in the U.S or immigrated to the country before age 6, and Group 3 also included participants who were born in the US However, it could also include participants whose at least one of whose parents fit the definition of the participants in Group 2. One of her specific research questions was the exploration of tense-mood-aspect across the Spanish continuum. Some of the most intriguing findings focus on the “simplification” and “loss” of tense-mood-aspect morphology in the Spanish spoken by the Spanish English bilingual adults in Los Angeles, which she considers an example of “Spanish language attrition”. Due to the fact that the functions of Spanish verb morphology are often complex, Silva-Corvalán decided to indicate three levels of “simplification” and “loss” depending on its linguistic context. The first level refers to the obligatory syntactic context, including instances that required a different verb form; the second level refers to discourse-pragmatic contexts, and the third level refers to optional contexts.

Her qualitative analysis showed that most instances of the past tense fit under the “simplification” or “loss” of obligatory syntactic contexts. Silva-Corvalán found that a simplification of the Preterite/Imperfect opposition is present. An example of this is shown in (8) where the speaker makes use of the Imperfect in an instance where the Preterite is “required”, based on the rest of the information in the sentence. The speaker

mentions *un accidente* “an accident” which indicated a perfective event that had a specific end. The author additionally pointed out that this type of rule was never “broken” by the first generation speakers in Group 1. Instead, the rule is found frequently “violated” by the heritage speakers found in Group 2 and Group 3. This phenomenon found in Mexican-American speakers of East Los Angeles demonstrates the alternation between morphosyntactic forms that do not share the same aspect, therefore the author argued that they don’t share the same meaning.

(8) “Iba a ser profesional, pero creo que **tenía** un accidente.”

‘He was going to become professional, but I think he *had* an accident.’

(Silva-Corvalán, 1994, pp. 24)

Silva-Corvalán’s study aimed to investigate whether the Spanish English bilinguals in these specific contact situations developed strategies in order to lighten the cognitive load of both of their linguistic systems. Her findings of what she considers “simplification” or “loss” of the Preterite/Imperfect opposition illustrate that these bilinguals make use of semantic extensions of the verb tenses. As previously stated, she shows that the first generation Spanish English bilinguals in Group 1 do not illustrate “simplification” of the Preterite and Imperfect while Group 2 and Group 3, the groups which include heritage speakers, do. It is important to point out that proficiency level was a present factor when showing signs of this “simplification”.

Participants on the lowest side of the proficiency continuum of Spanish were the participants that typically used this “simplification”, while the speakers in the high end of

the proficiency continuum retained the Preterite and Imperfect distinction. The pattern of “simplification” was a preference to refer to the past tense of stative verbs with the Imperfect, which suggests that a neutralization between Preterite and Imperfect distinction has occurred within a specific set of verbs. These verbs are *estaba* ‘was’, *era* ‘was’, *tenía* ‘had’ and *sabía* ‘knew’. On the other hand, a preference for using the Preterite form with non-stative verbs is also present. Another interesting finding is that speakers do use the Preterite form *fue* ‘went’ which in Spanish is a homophonous form to *fue* ‘was’. Moreover, the *fue* ‘was’ form to refer to the verb ‘to be’ was not found at all within the speech of the participants that had neutralized both types of verbs. *Fue* only works as ‘I went’ and *iba* only works as a verbal periphrasis ‘going to +’ not its other Imperfect form to signal ‘I went’. This finding supports a claim of “simplification” where stative verbs are preferred with the Imperfect and non-stative verbs are preferred with the Preterite. Additionally, it suggests that to these speakers, it is important to signal the past tense. However, the aspectual features are not as prominent.

In a similar context of language contact, Zentella’s (1997) seminal book gives us an insight into the social and linguistic realities of 20 Puerto Rican families in El Bloque, a low-income neighborhood in New York. She follows the families for 13 years. This longitudinal anthropological and sociolinguistic work enlightens us with the trajectory of these bilingual speakers as they head towards both language maintenance and language loss. One of the linguistic peculiarities Zentella describes in her book is the Preterite boundary in the bilingual continuum. Similarly to Silva-Corvalán’s findings, some of the Spanish English bilingual speakers preferred to use the Imperfect with stative verbs in perfective contexts. To illustrate this example better, (9) shows an excerpt from

Zentella's book where we can clearly observe that this speaker is using the verb *era* 'was' in its Imperfect aspect, instead of using the verb *fui* 'was' in the Preterite. Given that the verb *pagué* 'paid' signals that the action was completed and seen as a whole, *fui* 'was' would have indicated this completed action.

(9) “Yo *era* la que pagué por eso.”

‘I was the one who paid for it’

(Zentella, 1997, pp. 187)

The works discussed illustrate how in certain cases, the Imperfect and the Preterite neutralize to solely reflect the past tense. As previously mentioned, the simple past in English assumes a perfective aspect. Therefore, a direct transfer to Spanish would most likely increase the use of the Preterite and its perfective aspect. Both studies highlighted that the Imperfect/Preterite opposition is a lot weaker with stative verbs and that it is mainly in these cases where the Imperfect morphology is used to cue the past tense. Non-stative verbs on the other hand appear to for the most part to retain a higher Imperfect/Preterite aspectual opposition.

Cross-generational studies

As seen in the previous section, the highlighted non-canonical uses of the Preterite and the Imperfect were produced by speakers of second and third generation, or in a more general sense speakers who were either born in the US or moved to the US at an early age. Studies that have observed generation as a relevant factor have been specifically interested in Spanish proficiency in the second generation Heritage Speakers

(HS) (Montrul, 2014). Their proficiency has been analyzed from its phonological perception and pronunciation (e.g., Kim, 2012), as well as from a morphosyntactic level (e.g.; Alarcón, 2011; Bruhn de Garavito & Valenzuela, 2006; Montrul, 2004; Montrul & Bowles, 2009; Montrul et al., 2013; Otheguy and Zentella, 2012; Pascual y Cabo, 2013; Silva-Corvalán, 1994). The studies focused on morphosyntactic variables have found that the grammatical systems of the heritage speakers vary across generations.

The studies that have analyzed the tense and aspect Spanish system have typically suggested that their cross-generationally data shows patterns of “reduction” and “simplification” (e.g., Lynch, 1999, Martínez Mira, 2009, Mikulski, 2010, Montrul, 2007, Montrul, 2009, Montrul and Perpiñán, 2011, Silva-Corvalán, 1994). Terms that are commonly associated with the speech of heritage speakers within the US. The process or strategy of simplification has been described as “the higher frequency of use of a form X in context Y (i.e., generalization) at the expense of a form Z, usually in competition with and semantically closely related to X” (Silva-Corvalán, 1994, pp. 3). Most of the studies listed under studies that have analyzed the Spanish tense and aspect across generations were focused on the use of the Subjunctive and Indicative forms. However, aside from finding aspects of “simplification” some studies have also found similarities when comparing to monolingual Spanish speakers (Martínez Mira, 2009). Generally speaking, most studies analyzing linguistic variables in heritage speakers across generations have analyzed these variables with its canonical uses as the base or have made comparisons with monolingual speakers which have then lead to point out strategies such as the simplification strategy. However, this study aims to let the data speak for itself without preconceived notions of the variable at hand and without comparisons to monolingual

varieties. Instead, the analysis in this dissertation searches for the Spanish spoken by the three generations being observed to provide us with patterns and parameters to how the Preterite and Imperfect are being used.

In sum, previous studies related to both the linguistic variables of interest; namely, the use of the Preterite and the Imperfect, as well as how these variables have been studied in both monolingual and bilingual contexts and across generations. As summarized, most of the work done on the Preterite and Imperfect either assumes a concrete and universal understanding of the use of the Preterite and the Imperfect and seeks to analyze which speaker gets it “right” or wrong” or utilize purely experimental methodology to determine usage. This dissertation aims to shed light on the grammar of Spanish English bilinguals in Arizona with respect to the Preterite and Imperfect use without considering any of these traits as errors or mistakes. Instead, one of the goals is to contribute to the quantitative studies that have shed light on linguistic characteristics commonly used in Spanish in the US

Summary

The purpose of this chapter was to contextualize and identify relevant research focused on the Spanish Preterite and the Imperfect among US Spanish speakers. As the review showed, the trajectory that had led to the study of morphosyntactic variables such as the uses of the Preterite and the Imperfect has now led to an increase in variationist studies focusing on these types of linguistic variables observed in different varieties of a language in order to show how language is being used and how language changes over time. This chapter also showed that few studies have focused on utilizing corpus data

with the goal of observing the uses of the Preterite and Imperfect with either monolingual or bilingual speakers. Perhaps the preference to observe both of these past forms through experimental methodologies was due to the assumption that with the regard to the Preterite and the Imperfect no variation exists. However, more recent studies have shown that variation is present in both forms in both monolingual and bilingual speech although the research is limited.

This chapter also presented important works that have observed generation as a factor and discussed other relevant concepts that intertwine with language use between generations such as language maintenance and language loss. These concepts are important to acknowledge when dealing with speakers from different generations. More specifically it highlighted some of the difficulties that different generation of speakers go through in an anti-bilingual setting such as the US. Especially those of the later generations.

The next chapter will present the methodology employed to address my research questions, including where the data is derived from, and the linguistic and extra linguistic factors being observed.

CHAPTER 3

METHODOLOGY

The current chapter begins by reminding the reader of the rationale of this dissertation. This rationale was compiled by previous research and presents the gaps in the previously discussed literature which shaped this dissertation's research questions. Secondly, I provide the envelope of variation that was followed prior to the analysis of the data. Lastly, this chapter discusses further methodology steps utilized in this dissertation which include where the data analyzed is derived from, an explanation of both the linguistic and extra-linguistic predictors that are being coded in this dissertation, and a description of how the data was statistically analyzed.

The rationale

The goal of this dissertation is to analyze the use of the Preterite and the Imperfect in three generations of Spanish and English bilingual speakers in Arizona by using corpus data. This goal aims to cover two main gaps in the literature. First, most of the previous research that has studied these two past forms approach the variables from a generativist perspective, an approach that does not typically consider variation (see Chapter 2). Therefore, a sociolinguistic approach to the use of the Preterite and Imperfect is necessary. These two approaches are distinct from each other. A generative approach, while focusing on "correct" forms prescribed by standardized grammars, aims to obtain accuracy related to perception and/or production, which can sometimes ignore the possibility of variation that may be present with the linguistic forms being studied; in this case, the Preterite and the Imperfect. Therefore, any form that does not follow traditional,

standardized grammars would be observed as wrong. A sociolinguistic and variationist approach conducts a more bottom-up analysis that focuses on how language is being used. From this perspective, the notion that languages are constantly changing and evolving allows for the presence of variation. In addition, a variationist perspective allows for the analysis to focus on different internal and external predictors that favor the linguistic variable(s) being observed rather than determining if a variable is correct or wrong.

Second, recent research has shown that the dichotomy of the Preterite and the Imperfect is not as clear as traditional Spanish grammars suggest (e.g., Delgado-Díaz, 2014; Delgado-Díaz, 2018), and that it is still taught in many of Spanish language classrooms. In addition, there is still a lack of empirical research on these past forms in natural speech. Therefore, empirical research that is able to demonstrate what favors the use of one past form over the other based on corpus data is also sorely needed. Lastly, while a few studies such as Silva-Corvalán (1994, 2003) and Zentella (1997) have highlighted the aspectual “overlaps” regarding the Preterite and the Imperfect in their qualitative work, those parameters have not been studied using quantitative methods.

A further goal of this study is to understand these two past forms in the speech of Spanish and English bilinguals in Arizona in order to obtain insights as well as a first glance on whether this use can be considered a characteristic of this variety of Spanish in the US and if the uses found in the data are consequence of an internal change due to the intensive language contact found in Arizona. As mentioned in Chapter 1, the immense presence of Spanish, bilingualism and language contact between English and Spanish not only in Arizona but the entire Southwest makes the presence and varieties of Spanish

found a special case when compared to the rest of the country. In addition, much more is known about other areas of the southwest such as California and New Mexico, while there is still much to be discovered about the Spanish use in Arizona.

Research questions

In order to achieve these goals, the following research questions were developed:

1. What is the distribution between the Preterite and the Imperfect in the Spanish of the Spanish and English bilinguals in Arizona? What are linguistic and extra-linguistic predictors of the choice between the Preterite and Imperfect?
2. Does the speaker generation influence the distributional patterns found in the first research question?
3. How do these results resemble previous studies conducted with monolingual and bilingual speakers?
4. What do these results tell us about this particular variety of Spanish in the U.S?

The data

A compilation of two corpora from Arizona was used for the analysis in this dissertation. The corpora used were the Corpus del Español en Phoenix Arizona (CEPA) and the Corpus del Español en el Sur de Arizona (CESA). To date, these are the only two corpora that exist in the state of Arizona and both of the corpora follow a sociolinguistic methodology.

CESA's first and current Principal Investigator (PI) is Dr. Ana M. Carvalho. This corpus provides transcriptions and audio as well as background information from interviews conducted with bilingual speakers in Arizona from different generations of speakers. These interviews were collected and transcribed by undergraduate and graduate students in the Tucson area of Arizona under the supervision of the PI. The interviews lasted approximately 60 minutes and the interviewees were asked to discuss themselves and their community. The interviews were conducted wherever the participant preferred but usually in their home or in a public space. The interviews were conducted in Spanish, although participants could switch to English whenever they wanted to. Participants were interviewed and digitally recorded in individual sessions that aimed at eliciting spontaneous speech. The project is IRB-protected, and all measures are taken to assure the participants' anonymity (Carvalho, 2012-).

CEPA's first and current PI is Dr. Álvaro Cerrón-Palomino. This corpus provides some background information and transcriptions to interviews conducted with bilingual speakers around Phoenix Arizona. These interviews were collected and transcribed by graduate students and ranged from 30 to 60 minutes. Similarly to CESA, the sociolinguistic interviews focused on asking about the volunteers and their community. The author of this dissertation was one of the graduate students that conducted interviews for the CEPA corpus. This project is also IRB-protected, and measures were taken to assure the anonymity of the participants.

All interviews followed the Labovian interview procedure. Participants were asked different topics in order to elicit different tenses and styles of speech (Labov, 1973). The sociolinguistic interview has typically been the most popular method to obtain

data. Labov has defined the sociolinguistic interview as a “well-developed strategy” that is defined by a number of goals (Labov, 1984, pp. 32). The interview is typically carried out using modules that follow an interview schedule beginning with general questions and moving towards more personal questions. Broadly speaking the main goal of a sociolinguistic interview is to elicit vernacular speech from the participants, meaning the most natural and uncontrolled speech possible. Vernacular speech gives researchers a real sense of how that person’s speech reflects that of their community. Importantly, the researcher is not asking questions to get information but instead spontaneous speech (Tagliamonte, 2006). All interviews ranged from 30 minutes to an hour-long.

A total of 30 sociolinguistic interviews were selected from the corpora. Table 1 describes the participants utilized in this study. These interviews were stratified according to predictors. Therefore, speakers that fit a certain generation and sex were picked in order to have a complete equivalent pool of participants. To be able to account for an analysis of the Preterite and Imperfect use of the Spanish English bilinguals in Arizona, the interviewees selected for this study had lived in Arizona for at least 5 years. All participants were at least 18 years of age or older and represented three different generations; a more detailed description of the generations is provided in the external predictors section below. The 30 interviews are divided into 10 that represent the first generation, 10 that represent the second generation and 10 that represent the third generation. In addition, the 10 participants in each generation were composed of five females and five males, except for the third generation that has six females and four males. The generation factor is described in more detail in the extra linguistic factors section of this chapter.

Table 1. Participants background information

Participant	Age⁴	City	Years in city⁵	Generation	Place of birth	Sex
1 [CESA074]	54	Tucson	12	1st	Sonora	F
2 [CESA075]	57	Tucson	22	1st	Sonora	F
3 [CESA069]	28	Tucson	17	1st	Sonora	F
4 [CESA059]	25	Tucson	12	1st	Sonora	M
5 [CESA054]	68	Tucson	30	1st	Sonora	M
6 [CESA039]	54	Tucson	40	1st	Sonora	F
6 [CEPASI]	33	Phoenix	5	1st	Guadalajara	M
7 [CESA032]	25	Tucson	23	1st	Sonora	F
8 [CESA021]	25	Tucson	22	1st	Sonora	M
9 [CESA002]	31	Tucson	9	1st	Sonora	M
10 [CESA049]	22	Tucson	22	2nd	Tucson	M
11 [CESA043]	18	Tucson	18	2nd	Tucson	F
12 [CESA042]	32	Tucson	32	2nd	Tucson	F
13 [CESA044]	21	Tucson	16	2nd	California	M
14 [CESA076]	29	Tucson	29	3rd	Tucson	F
15 [CESA067]	25	Tucson	25	3rd	Tucson	M
16 [CESA050]	21-29	Tucson	21	3rd	Tucson	F
17 [CESA036]	27	Tucson	35	3rd	Morenci	M
18 [CESA027]	20	Tucson	20	3rd	Tucson	F
19 [CESA024]	21	Tucson	21	3rd	Tucson	F
20 [CESA045]	29	Tucson	6	3rd	Texas	F
21 [CESA006]	43	Tucson	41	3rd	El Paso	M
22[CEPAMC]	20	Phoenix	14	1st	Mexico City	F
23 [CEPANS]	28	Phoenix	28	2nd	Juarez	F
24 [CEPAHM]	23	Phoenix	20	2nd	Phoenix	M
25 [CEPAAI]	28	Phoenix	10	2nd	Nogales,AZ	M
26 CESA022	21	Tucson	21	2nd	Tucson	F
27 CESA018	25	Tucson	25	2nd	Tucson	F

⁴ Age at time of interview

⁵ Years lived in city at time of interview

28 CESA038	33	Tucson	28	2nd	Tucson	M
29 CEPAVAE10M3 BE	NA	Phoenix	66	3rd	Mesa, AZ	M
30 CESA073	22	Tucson	22	3rd	Tucson	F

As Table 1 shows, the participants' ages ranged from 18 to 68 at the time of their interview. There are also a couple of participants whose ages were not explicitly provided. For example, participant 16 did not provide her exact age but thanks to context given in the interview and the fact that she had lived in Tucson for 21 years at the time of the interview, the researcher assumed the interviewee's age ranged between 21-29. As a result, when coding the age for this participant, 25 was chosen as this was a number in between the 21-29 range. The other participant whose age was not explicitly provided was participant 29. However, given the context of the interview, and information that showed the participant had lived in Arizona for 66 years, the information suggested that the participant was that age at the time of interview.

The information provided in Table 1 also suggests that most of the speakers were either born in the southwest of the US or in the northern area of Mexico. This shows that most of the participants were in close proximity to the place they were born or the place where they have family members residing. Overall, both corpora contained more speakers of first and second generation. It was more difficult to gather interviews from third generation speakers. This could have been due to the amount of third generation speakers who were willing to be interviewed. As mentioned in Chapters 1 and 2, many third and later generation speakers have a harder time with their linguistic identity and confidence.

Envelope of variation

Given that one of the main research questions for this study aim to discover what the constraints of both of these past variables are, the current envelope of variation solely focuses on the instances that stood out when coding. To begin, the variable to be studied will be described as follows: If a participant rephrases themselves such as in (10), where the participant initially uses the verb in the Preterite form but then corrects themselves and produces the same verb in the Imperfect tense, solely the second verb produced was included in this study⁶.

(10) Yo no recuerdo, nunca participé, **participaba** en esas cosas.

(PhoeHM_116)

If the use of the verb was not clear, these tokens were not included. An example is when a verb was conjugated in either the Preterite or Imperfect, but the context of the use was not very clear. An example of an instance that was not analyzed is shown in (11).

(11) shows two instances that were accounted for in the data: *decía* ‘I said’ and *hacía* ‘I did’. However, the verb *me quedé* was not accounted for due to lack of clarity. Similarly, the instance of *hacimos* in (12) is ambiguous; it is not clear if the speaker meant *hacíamos* or *hicimos*. Instances such as those were not included in the data. Another example where it was difficult to gather what exactly was meant is shown in (13) where

⁶ These examples will be kept in a different document for future research

the participant says *tenían* drunk. While this could have easily been coded as a third person plural, the meaning of this part of the sentence was not clear, therefore, it was not coded. As previously mentioned, many of these examples will be interesting starting points for future research.

- (11) Que no tengan miedo y precisamente lo que yo **decía** no lo **hacía** yo mismo
¿Me entiendes? *So* yo **me quedé** [ininteligible] por qué estoy diciéndoles a ellos
que siguen los sueños, trabajen duro para lograrlo y haz lo que tú quieras.

(CEPAAI_255)

- (12) Y cuando **hacimos**, cosas ellos me digan “Oh estabas correcto”

(CESA067_411)

- (13) Pero como los customers a veces you know como estamos aquí en universidad
tenían drunk y todo eso antes de los juegos de football eso era un poquito
difícil...

(CESA050_577)

There were instances, especially within the third generation, where the researcher had to interpret what was being said based on the context. If context did not provide additional information, the instance was excluded. However, in an example such as (14) where the participant conjugated *nacer* with the third person *nació* to mean *nací*, this would have been coded as Preterite first person singular. The context, as well as the

interviewer's interruption where they say "*oh, okay, fue antes que nacieras*" and this is followed by the participant's *sí* gives the necessary information to code it as such. No other forms of the past were coded, all false starts and any possible interruptions were also excluded.

(14) ¿Él? /Mhm/ No sé, no, fue antes que, que **nació** y /oh, okey okey, fue antes que nacieras/ sí...

(CESA076_322)

Factor groups

This section begins describing the factor groups that will be analyzed in this study as well as possible predictions. More specifically, the researcher has taken into account the relevant predictors in studies previously discussed in the literature review. These predictors will hopefully inform us whether speakers favor the use of the Imperfect or Preterite and how this may be different depending on language dominance.

Internal predictors

Aspectual Interpretations

The aspectual interpretations coded for in this study were the perfective and imperfective. The imperfective categories included were habitual and non-habitual. Additionally, a category for instances that were inconclusive/indeterminate⁷ was created.

⁷ Inconclusive to the researcher, similarly they could be instances where the context surrounding the tokens did not explicitly give perfective or imperfective interpretations

All tokens were coded based on the context surrounding the variables being observed. Discourse context, as well as surrounding adverbs, were useful for attributing an aspectual reference. Aspectual meaning is an intractable concept, and as such, the coder can only make assumptions, which are influenced by the researcher's subjectivity (Schwenter & Torres-Cacollous, 2008).

As presented in the introduction, the perfective aspect is seen as a single event that typically demonstrates an endpoint or culmination, such as (15) where the verb *nací* "I was born" refers to an act that was done once and had a relatively specific end. A few different functions fall under the Imperfect aspect, which includes habitual and non-habitual instances. The instances coded habitual were the events that due to contextual information were repeated over an extended period of time such as (16). In this case *ayudaban* and *estaba* appear to be actions that occurred over a certain period of time when this participant was in school and would go visit his friend's house. While there are no specific adverbs that clarify this happened more than once the context surrounding the verbs shows that visiting their friend's house occurred more than once/occurred multiple times. Example, (17) presents an example of an imperfective non-habitual aspectual interpretation.

The instances coded under the non-habitual aspect fall under the imperfective aspectual range (see Chapter 1). These instances can include ongoing actions in a specific temporal frame. Some instances do not necessarily fall under progressive; however, they do not fall under habitual instances either. (17) shows an example of this instance. This sample shows the verb phrase *andaban estresados*, an instance that is presenting an imperfective aspectual function given that there is no exact end to the stress of the

participant's peers. This instance is not clearly a habitual aspect. Lastly, as previously indicated, in the case of ambiguity due to either the absence of discursive or contextual cues, those instances went into an 'inconclusive' category, an example of which is (18). I hypothesized that most of the data would follow the pattern showing the Preterite verbs with a perfective aspect and the Imperfect showing an imperfective aspect. However, I also hypothesized that instances that didn't follow this pattern will be found in the Imperfect spectrum.

Perfective:

- (15) **Nací** en Juárez México y vine... nos movimos para acá cuando tenía como un año.

(PhoeCoNS_2)

Imperfect:

Habitual:

- (16) Los papás también nos **ayudaban** mucho con la tarea cuando yo **estaba** ahí, when in my house, todos nosotros estábamos solos así que mis papás nunca nos podían hablar...ayudar si teníamos preguntas con la tarea y todo eso.

(PhoeCoNS_35)

Non-habitual:

- (17) Pero no era difícil, como tenía otros amigos, que **andaban** estresados, frustrados, porque tenían que um (...) /subir nota o.../ Sí

(CESA032_914)

Inconclusive:

(18) ...y con ellos- con ellos conviví, ellos me **ayudaron** y yo les ayudaba, e:h, con la habilidad de los planos, de dibujar,

(CESA074_1177)

Verb Lemma and Lexical Frequency

In order to obtain a deeper understanding of what verbs may favor one variant over the other, and to be able to analyze the verb effect on the variables, the verb lemma (verb in the infinitive) is also coded and was analyzed as a random variable. In their recent study, Orozco and Hurtado (2020) included verb lemma analysis with the goal of obtaining a more detailed analysis of the effects of the verb regarding the subject pronoun expression variable. Their results showed that all verbs within a given category such as the lexical semantics do not promote the variable being studied. For this reason, it is necessary to see what part of variance is due to the verb lemma and what is due to the fixed effects, which often correlate within groupings such as verb lemma. Considering that this could possibly occur with more variables, this predictor allows the researcher to examine the verbs past the semantic categories previously mentioned.

By coding the verb lemma, the researcher could analyze whether certain patterns occurred in less or more frequent verbs. This numeric factor was coded by the number of instances each verb lemma presented itself in the first generation. Theoretical proposals such as the usage-based framework have suggested that speakers store more detailed information with more frequent expression (Bybee 2001; Pierrehumbert 2001; Bybee and

Torres Cacoullos 2008). Therefore, taking the verb frequency into account, while exploratory, can shed some light on the verb and the connection with the frequency of use. As will be mentioned in the next section, the frequency factor and proportion of the Preterite factor were observed together in the analysis.

Proportion of instances of the Preterite over the Imperfect

It has been demonstrated that lexical representations of patterns are maintained through high frequency in situations of linguistic variation. Linguistic forms are more resistant to regularizing patterns because of these strong lexical representations (Bybee, 1985; Brown & Rivas, 2012). Therefore, the proportion of the use of the Preterite over the total occurrences for each verb lemma gathered from the first generation. This was done to determine if the use of certain verb lemma were preferred in one past tense over the other in the first generation to observe if the variation patterns are influenced by whether a verb lemma was preferred to be used in the Preterite or the Imperfect.

This factor was included to observe the possible interaction between the frequency of the verb and the proportion of use in the Preterite. The hypothesis behind this predictor falls under the usage-based framework. Therefore, if a verb is of low frequency and in the first generation prefers to use this verb with the Preterite 90% of the time and only 10% of the time uses it with the Imperfect, then it could be hypothesized that the third generation, who may receive less Spanish input, may only get to receive input of that verb in the Preterite due to its high proportion of use in the Preterite over the Imperfect. This factor was considered a numerical factor where the instances of each verb lemma in the Preterite were divided by the total instances of the very same verb lemma.

Clause type

This predictor focuses on identifying if the instances are part of the main clause or a subordinate clause. Analyzing whether the verbs occur in a main or subordinate clause will reflect the foreground versus background syntactic placement as it has been shown that foreground information is typically found in main clauses, while background information in subordinate clauses (Hopper & Thompson, 1979). Different types of clauses provide different types of information; for example, studies have shown that foreground information is more likely to appear more with the Preterite while the Imperfect presents background information (Bardovi-Harlig, 2000; Hopper, 1979; Klein & Von Stutterheim, 1987). However, hypotheses for this predictor are conflicting. Previous research has shown that the Imperfect may be preferred over the Preterite (e.g. Silva-Corvalán, 1994, 2003; Zentella, 1997) which would most likely occur in subordinate clauses. However, Bybee and Thompson (2002) mentioned that main clauses are the ones expected to be innovative first, followed by subordinate clauses. This is due to the main clauses being pragmatically richer than the subordinate clauses. (19) shows the use of the Preterite *me imaginé* “I imagined” in the main clause and the use of the Imperfect *no era* “it wasn’t” and *estaba* “it was” in the subordinate clause as traditionally expected. However, variation is hypothesized.

(19) Entonces **me imaginé** que no **era** esa pesadilla. Pues que **estaba** chistoso.

(PhoeCoHM_153)

Grammatical Person and Number

While not a significant predictor in previous research, all conjugated verbs in either the Imperfect or Preterite are coded depending on whether they had an overt or null subject. Given that the data is formed by interviews, it is most likely that most if not all pronouns appear. However, since this study only accounts for verbs conjugated in two different pasts and bearing in mind that the interviews include a lot of narrations as opposed to spontaneous conversations, I did not expect to find many of any instances conjugated in the second person, either singular or plural.

Furthermore, it may be relevant to point out that the first person has been linked to the subjectivity of the event, meaning that it is closely based on the speaker's internal beliefs and attitudes (Schwenter & Torres Cacoullós, 2008). Therefore, it could be hypothesized that the first person may favor the use of the Imperfect due to the subjectivity of the event. This is hypothesized due to that the Imperfect expresses the development of the event internally to the event.

External predictors

Generation

Using the questionnaires completed by the participants, as well as the information provided in the interviews, participants were divided into three generations, first, second and third following Silva-Corvalán (1994). A total of 30 participants were divided into these three groups. 10 participants formed the first generation group, meaning parents as well as themselves were born in Mexico. 10 participants formed the second generation group, where both parents were born in Mexico but the participants themselves were born

in the US. Lastly, 10 participants formed part of the third generation. This group consisted of participants who were born in the US along with one or both parents.

While this dissertation is a synchronic study, the generation predictor is applied for two reasons: 1) to better understand whether latter generations, who may have had more contact with English, use this variable differently than prior generations, especially the first generation and 2) to compare with previous studies that although did not analyze this variable may, they may have noticed some patterns with the second and third generation.

Previous research can be used to hypothesize possible outcomes. In the case of the generation predictor, Silva-Corvalán (1994, 2003) demonstrated that most cases where the Imperfect was used with perfective aspects were produced by second and third generation speakers. Similarly, Zentella found “curious cases of the Imperfect” among the second generation Puerto Rican children in New York. An example is provided in (20) where Paca, a second generation Puerto Rican girl in Puerto Rico says *yo era la que pagué por eso* “I was the one who paid for it.” Two instances of the past are present, the verb *pagué* ‘I paid’ tells us that the act of paying has culminated, however, the *era* ‘I was’ does not complement the perfective aspect. Therefore, in this research, I hypothesized that the instances that stand out will most likely be produced by the second and third generations. This could include similar instances to Silva-Corvalán and Zentella’s work such as the Imperfect being used with a perfective aspect.

(20) “Yo *era* la que **pagué** por eso.”

‘I was the one who paid for it’

Language Dominance

The categories used for the language dominance predictor are English dominant, balanced, and Spanish dominant. Both corpora incorporated a different but specific way of determining language dominance. The CESA corpus contained a questionnaire in which one section was dedicated to self-rating language competence questions in both of the languages the participants spoke. In order to ensure that all participants were Spanish English bilinguals, participants were asked to self-rate their language competence in both English and Spanish. Participants rated their competence based on how well they thought they speak, understand, read and write in both English and Spanish. The ratings provided are from 0-6, where 0 codes for *no muy bien* “not very good” and 6 *muy bien* “very good” (Appendix B). Since both aspects of data gathering will require participants to produce oral data, participants would need to be more than receptive bilinguals and demonstrate a level of language competence by agreeing to be part of an interview in Spanish. For the purposes of this study, the higher the language competence scores are in Spanish the higher the language dominance rating it receives. Therefore, a 0-1 score was considered English dominant, a 2-4 score was considered balanced and a 5-6 was considered Spanish dominant.

The CEPA corpus included a questionnaire with 12 bilingual dominance scale questions, following Dunn & Fox Tree (2009) (Appendix A). A representation of the bilingual dominance scale is shown in Figure 2. In contrast with the CESA corpus, the 12 questions form part of the scoring procedure. CEPA has the researcher classify the participants under English-dominant, Spanish-dominant, or balanced bilinguals.

Additionally, the scale provided in the questionnaire targets three main criteria important in gauging dominance: “percent of language use for both languages, age of acquisition and age of comfort for both languages and restructuring of language fluency due to changes in the linguistic environment” (Dunn & Fox Tree, 2009, pp. 273). The scoring procedures can be summarized by grouping participants with scores of -5 to 5 as balanced bilinguals. The scores below -5 reflected English dominance, while the scores higher than 5 reflected Spanish dominance.

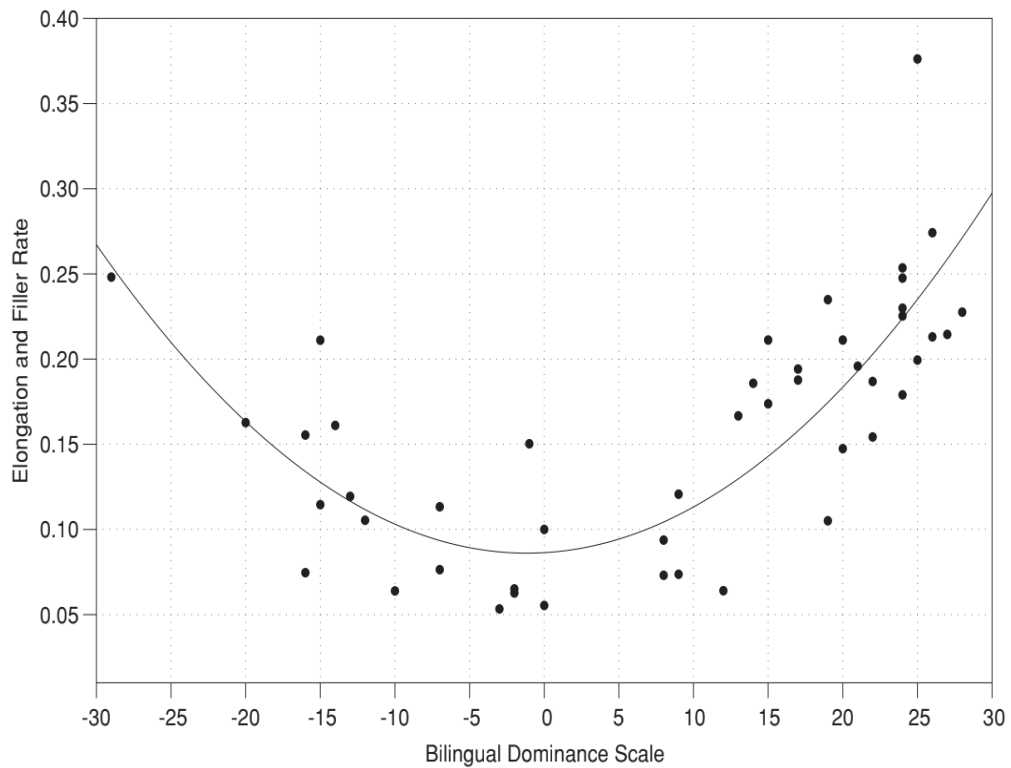


Figure 2. Elongation and filler rated by bilingual dominance scale (Dunn & Fox Tree, 2009, pp. 286)

Both generation and language dominance are being analyzed as different factors in order to observe what external factor related to their amount of language contact has a

stronger significance with how the bilingual speakers utilize the Preterite and Imperfect. In addition, the generation factor focuses on the speakers' approximate relationship in the US which can suggest how long they have been part of the language contact that is present in the US. On the other hand, the language dominance factor focuses on the speakers' self-beliefs of how strong they believe their dominance is in each language while still having the option of expressing they consider themselves "balanced" bilinguals. Lastly, these two factors were separated in order to highlight that language dominance does not always correlate with the generation of the speaker. For example, there are many second and third generation speakers that would consider themselves balanced bilinguals which goes against the typical understanding that second, third or further generations always go through language shift or language loss.

Age and Sex

From an experimental perspective, previous research has shown that the older bilingual children used a higher proportion of Preterite and that this overuse may stem from a transfer from English (e.g., Cuza, et al., 2013). From a corpus data perspective not much has been said with regards to age other than what is found in Delgado-Díaz (2014), where the results illustrate age as a significant predictor where the younger participants favor the use of the Preterite while the older participants favor the use of the Imperfect (Delgado-Díaz, 2014). Given that age has not shown much effect with regards to the use of the different past tenses, this predictor is mainly an exploratory predictor.

Similarly, sex has not been shown to have a strong effect on the use of the Imperfect or Preterite. Only one study was able to show that female native speakers used the Preterite more often than their male counterparts. Additionally, it showed that both

gender groups used the Imperfect and Present Perfect at similar rates (Geeslin & Gudmestad, 2010). Once again, similarly to the age predictor, sex is considered an exploratory predictor.

Analysis

All instances and predictors were coded into an excel spreadsheet for their respective variables. First, I highlighted the data distribution and any potential disparities between them, then provided the overall distribution of both the Preterite and Imperfect, followed by the distribution within each predictor category. Before the data were subjected to multivariate analysis, this stage disclosed categorical and variable contexts (Tagliamonte, 2014, pp. 300). Additionally, I show conditional trees that help visualize the trends that are occurring with the Preterite and the Imperfect. Next, I conduct a multivariate analysis using Rbrul (Johnson, 2019). This analysis included all of the linguistic and extra linguistic predictors with Participant and Verb Lemma as random factors. An analysis in Rbrul was completed in order to provide the canonical information variationist studies typically show which are the token numbers, the predictor weight, p-value and range and to establish a hierarchy of factors.

In addition, mixed-effects logistic regression was conducted in R (R Core Team, 2020), where predictors were imputed as either fixed or random predictors. First, the MASS package (Ripley, 2017) was utilized to run a stepwise logistic regression for model selection, after the best model was selected from/by the mixed-effects logistic regression. The full model included a two-way interaction between Generation and the Aspectual Function plus the predictors gathered from the stepwise regression which were

Grammatical person, Grounding, Sex, Language Dominance and Age. Participant and Verb Lemma were also included as random factors. Mixed-effects logistic regressions were conducted in RStudio (RStudio Team, 2020), using lme4 (Bates, Maechler, Bolker, & Walker, 2015) and optimx (Nash & Varadhan, 2011) packages. Figures to illustrate the data were produced using the package ggplot2 (Wickam, 2009).

A combination of three statistical models were conducted for the following reasons: The conditional trees were first illustrated in order to provide a general idea of the tendencies each the Imperfect and Preterite showed with regards to when they were favored in the data. Following this, a multivariate analysis was conducted in Rbrul where all linguistic and extra linguistic factors were included and the Preterite was selected as the application value. In this analysis verb lemma and participant were included as random factors. While the conditional trees conducted in RStudio provide general trends, the Rbrul multivariate analysis was able to provide not only statistical significance, but also canonical and relevant data typically provided in a variationist study such as the number of tokens, the predictor weight and p-value per predictor as well as a range value. Lastly, all significant interactions were run in RStudio by way of mixed-effects regressions. In this step, the interactions between the significant factors were further explored as well as graphed by way of ggplot2.

Summary

The dissertation's rationale, research questions, and information as to where the data analyzed was gathered from and the envelope of variation were presented in this chapter. In addition, I presented an overview of the internal and external factors analyzed

with respective examples taken from the data utilized in this study. Each internal and external factor was also presented with a hypothesis. Lastly, I explained how the data was coded and analyzed to demonstrate how the results were reached. The next chapter focuses on the results and discussion that were derived from the data analyzed.

CHAPTER 4

RESULTS AND DISCUSSION

This chapter presents the dissertation's results. The goal of this dissertation was to provide a variationist analysis of the uses of the Preterite and the Imperfect in three generations of Spanish and English bilingual speakers in Arizona. The descriptive and inferential statistics of the Preterite and the Imperfect are presented in steps. Each subsection answers one of the research questions. The first research question is answered by presenting the distribution of both the Preterite and Imperfect tenses within the three generations to get a better understanding of how often these forms are used by the participants. In addition, the chapter presents the overall distribution of the predictors as well as the significant predictors that stand out within the data. The significant predictors will be corroborated by excerpts that demonstrate their use. Trends that are present in the data but not statistically significant are also presented with examples to give the reader a clearer understanding of these trends. Second, I answer the research question which presents the main differences in the results divided by generation. Third, I compare the current results to previous studies that have observed these two variables in both monolingual and bilingual speakers. Finally, the last research question is answered by providing insights into what the results suggest about this particular Spanish variety in the US along with some pedagogical implications.

Overall Preterite versus Imperfect use

The first research question was: “What is the distribution between the Preterite and the Imperfect in the Spanish of the Spanish and English bilinguals in Arizona? What are linguistic and extra-linguistic factors that predict the use of the Preterite and Imperfect?” A total of 2520 tokens were analyzed for this study. The overall frequency of the Imperfect versus the Preterite in the Spanish English bilinguals in Arizona appears in Table 2. This table shows that while there were more instances of the Preterite the difference is negligible; both the Preterite and the Imperfect were similarly present in the data.

Table 2. Overall frequencies of the Preterite and Imperfect

Grammatical Aspect	%	N
Imperfect	46.3 %	1166
Preterite	53.7 %	1354

Aspectual interpretations

Table 3 provides a deeper observation of the distribution of the Preterite and Imperfect within the independent linguistic factors. As was expected, the canonical distribution of the imperfective aspectual interpretations was mainly used with the Imperfect (93.9%) and the perfective aspectual interpretation with the Preterite (96.9%). However, it is interesting that the Preterite expressed with an imperfective aspectual interpretation is the main trend in the instances where the canonical interpretations are

not used. In addition, Imperfect verbs exhibit an “indeterminate” aspect (68.1%) more frequently than Preterite verbs (31.9%).

Clause Type

In terms of clause type, which was analyzed in order to obtain foreground and background information according to its syntactic placement, the data shows higher use of the Preterite (57.6%) in main clauses and a higher presence of the Imperfect (52.1%) in subordinate clauses. These results match what was expected and proposed by previous studies. As the table shows, the Preterite is more present in main clauses which typically provide foreground information (Bardovi-Harlig, 2000; Hopper, 1979; Stutterheim & Klein, 1987). Likewise, the Imperfect is mostly found in subordinate clauses which typically bear background information.

Table 3. Frequencies of the Imperfect and Preterite and internal predictors

	Imperfect		Preterite	
	%	N	%	N
Aspectual Interpretations				
Imperfective	93.9%	982	6.1%	64
Perfective	3.1%	39	96.9%	1222
Indeterminate	68.1%	145	31.9%	68
Clause Type				
Main	42.4%	641	57.6%	871
Subordinate	52.1%	525	47.9%	483

Grammatical Person				
1s	38.5%	397	61.5%	633
1p	48.2%	123	51.8%	132
2s	73.9%	17	26.1%	6
3s	49.5%	446	50.5%	455
3p	58.8%	183	41.2%	128
Verb Lemma				
<i>ser</i>	73.6%	246	26.3%	88
<i>tener</i>	77.6%	153	22.3%	44
<i>estar</i>	82.8%	150	17.1%	31
<i>ir</i>	26.6%	37	73.3%	102
<i>decir</i>	28.8%	34	71.1%	84
<i>hacer</i>	25.6%	20	74.3%	58
<i>empezar</i>	5.5%	4	94.4%	68
<i>vivir</i>	55.9%	33	44.0%	26
<i>venir</i>	11.1%	6	88.8%	48
<i>haber</i>	93.6%	44	6.3%	3

Grammatical person

The grammatical person factor was chosen with the hypothesis of the Imperfect being linked to the subjectivity of the event (Schwenter & Torres-Cacoullos, 2008). According to this hypothesis, I expected the Imperfect to be utilized mostly with the first person as the Imperfect as this grammatical feature expresses the development of the event internally to the event. Instead, the data show more instances of both the first-person singular (61.5%) and first-person plural to be used with the Preterite (51.8%). In

contrast, the distribution of the data is more balanced between the Preterite (50.5%) and the Imperfect (49.5%) in the third-person singular. In addition, the Imperfect (58.8%) is more common with the third-person plural than the Preterite (41.2%). The instances of the second person singular were originally going to be taken out of the data due to the relatively small number of tokens. However, they were kept in the analysis because of a compelling pattern that was observed in the descriptive statistics and examples.

Interestingly, most of the instances of the second person singular were used with the Imperfect (73.9%) and were not very frequent with the Preterite (26.1%). The most notable uses of the second person singular instances is that in most cases, it was the indefinite grammatical person: the referent of its subject is equivalent to the indefinite pronouns *uno, una* o *une* as seen in (21). Both *sabías* and *quedabas*, although presented in the second person singular, have a subject whose referent is indefinite. This use of the second person singular and the Imperfect to encode an indefinite meaning was not present with the Preterite. This pattern supports the canonical notion that the Preterite, which is typically connected to the perfective aspectual function may have a more solidified semantic meaning to the speakers. This would mean that when the speakers utilize the Preterite, there is an assurance of who experiences the state or event in that perfective aspect. On the other hand, the use of the Imperfect and in this instance its relation to an imperfective aspectual function conveys a more open semantic meaning that can extend to an indefinite person and indefinite state or action.

(21) Había mucha acción dentro de la película, y era algo en donde no **sabías** exactamente lo que iba a pasar um... siguiente. Entonces te **quedabas** en suspenso y se me hacía muy interesante la película.

(CESA044_2015,2017)

Verb Lemma

The last predictor presented in Table 3 is verb lemma. This predictor was included to observe and analyze the verb's frequency effect on both the Preterite and the Imperfect. The verb lemma factor was accompanied by a column that determined the proportion of the verb utilized in the Preterite in the first generation for comparison. This will be demonstrated later in the chapter. Due to the long list of verbs displayed in the data, only the ten most common verbs in term of frequency were included in Table 3 (see rest of the list in Appendix A). As the table shows, *ser*, *tener*, *estar*, *ir*, *decir*, *hacer*, *empezar*, *vivir*, *venir*, and *haber* are the ten most utilized verbs in the data. One notable pattern is that out of this list, the verbs that are mainly utilized with the Imperfect are stative or copula verbs such as *ser* (73.6%), *tener* (77.6%), *estar* (82.8%), and *haber* (93.6%) and the verbs mainly utilized in the Preterite are the non-stative verbs such as *ir* (73.3%), *decir* (71.1%), *hacer* (74.3%), *empezar* (94.4%) and *venir* (88.8%). An analysis of how these verbs behave with the Preterite and Imperfect is presented later in the chapter.

The Preterite

After establishing the overall distribution of the Preterite and the Imperfect and the notable patterns as they interact with the internal predictors, the next step was to determine what predictors favor the Preterite and the Imperfect. To obtain a visual representation, a conditional tree was performed with the statistical software R. The conditional tree can also show significance as well as constraint hierarchy (Delgado-Díaz, 2018; Scrivner & Díaz-Campos, 2016; Tagliamonte, 2012). However, in this chapter I am presenting conditional trees only to help visualize the trends that are occurring with the Preterite and the Imperfect.

Figure 3 illustrates that the most relevant predictor of the Preterite is aspectual function, and it interacts with the frequency of the verb, generation, and grammatical person. The conditional tree shows that the perfective aspectual function differs from the imperfective and indeterminate aspectual function. One of the main distinctions is that perfective instances highly favor the use of the Preterite as previously noted in Table 3.

As Figure 3 shows in Node 4 the Preterite is highly favored when the grammatical person is first-person singular, second-person singular, third-person singular and third-person plural, and when these grammatical persons are utilized in a verb that has a frequency number of 37 or less. An example of this can be seen in (22) where the verb *conocer* has a frequency number of 13 and is presented in the first-person singular. Another instance where the Preterite is favored is when the verb has a frequency number of 37 or less and first-person plural is the grammatical person as Node 5 shows. An example of this instance is presented in (23) where the verb *venir* which has a frequency number of 18 is conjugated with the first-person plural.

Figure 3 shows that while all these instances favor the use of the Preterite, the instances highlighted by Node 4 are more numerous. In addition, as Node 6 shows, the verbs with a frequency of 37 or more also tend to appear with the Preterite; however, this is not as favored as instances presented in Node 5. An example of this is shown in (24) where the verb *ser* with a frequency of 117 is presented with the Preterite. These first representations show a feature that was hypothesized from the beginning which is the canonical and most frequent use for the Preterite.

SHL instructors should take grammar lessons from the classroom as supplementary tools rather than norms to be applied in the classroom. While grammar points and examples that textbooks demonstrate are common uses and found in different varieties of Spanish, many non canonical uses are sometimes still presented as ungrammatical. As Beaudrie, et al. (2014) recommended, instructors should first establish what students already know (pp. 162). After having this information, instructors should focus on presenting examples that highlight the complexity of variation by providing summaries of variationist studies that present linguistic variation and showing the students' authentic data.

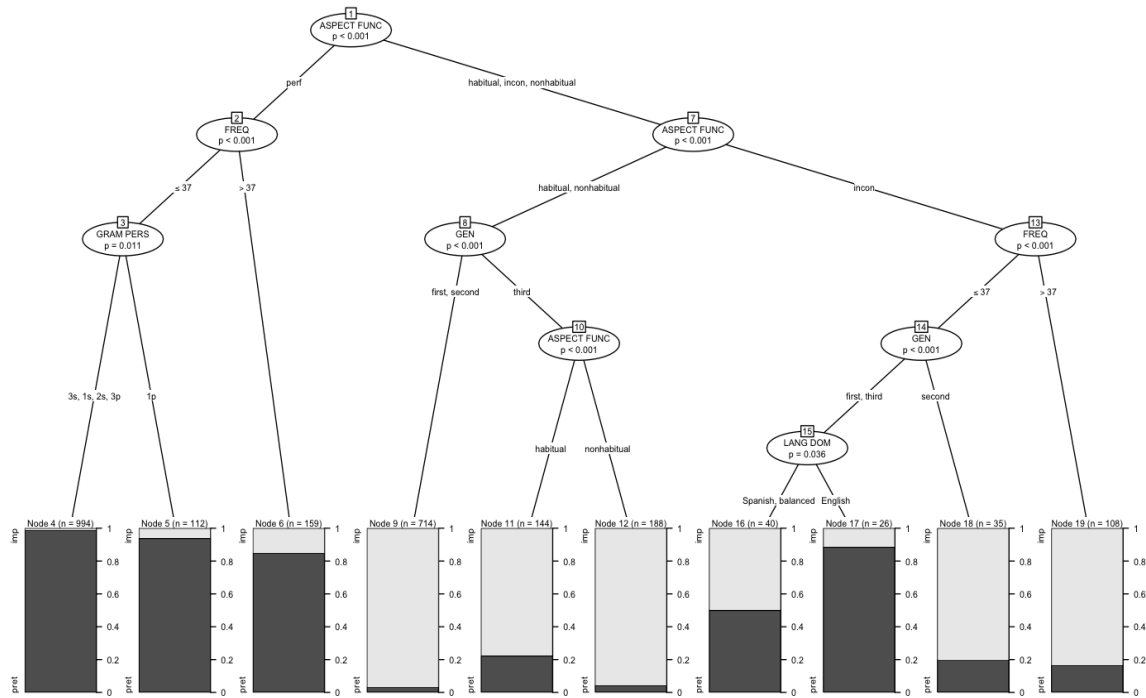


Figure 3. Conditional tree of the Preterite

(22) So, yo cuando trabajé con el Art institute de Phoenix. umm yo **conocí** un diseño gráfico allí que se llama Eric Pierce.

(CEPAAI_208)

(23) Um, no es- primero nos fuimos al pueblo, allá con mi familia, y nos **vinimos** porque murió una de mis hermanas /Mhm/. Nos **vinimos** y ya nos quedamos un tiempo allá, y ya después empezaba a crecer los niños y ya nos regresamos a Tucson.

(CESA074_483)

(24) No. Nunca tuvimos esa, y es raro porque esto cuando **vine** a oírlo fue aquí en los Estados Unidos.

(CESA039_1843)

Figure 3 shows that the presence of the Preterite is also favored when the imperfective habitual aspectual function is used by different generations, more specifically by the third generation as shown in Node 11. An example is illustrated in (25) with the same verb as the previous example. In this instance the participant was explaining the work hours his dad used to work when his parents were younger. The context around his story gives information that the times he went to work occurred more than once as they were part of his schedule, therefore although the participant gives this instance an imperfective habitual aspectual function, the grammatical form he chooses is the Preterite.

The right side of the tree focuses on the inconclusive instances. While these instances will not be deeply explored in this dissertation, I will present some examples that demonstrate why they are ambiguous/indeterminate with respect to an aspectual function at the end of this section. However, it appears that these instances were influenced by the generation and frequency of the verb as well as the language dominance predictor. When the verbs were of a frequency of 37 or less and produced by the first or third generation, they appeared more than with the second generation. In addition, as Nodes 16 and 17 show, the English dominant bilinguals preferred to use the Preterite a bit more than the Spanish dominant and balanced bilinguals. A similar pattern

to what occurs with the second generation is seen with the verbs with a frequency of 37 and over without any influence of a generation.

(25) Uh porque cuando eran chicos, mi mamá, más o menos mi mamá, no mi papá porque él **trabajó** como a las 3 de la mañana, hasta las-las 7:30 para ir a la escuela. Luego iba a la escuela. Luego iba al trabajo otra vez, pero mi mamá no.

(CESA067_393)

The Imperfect

This section uses another conditional tree to highlight and provide examples of the predictors of the Imperfect. As expected, many contexts that favor the use of the Imperfect are the opposite of the ones that favor the Preterite. Figure 4 shows that the most important predictor to favor the Imperfect is aspectual function, followed by frequency of the verb, generation, and language dominance. As Node 14 shows, when an instance is imperfective non habitual and produced by a third generation speaker the presence of the Imperfect is high. A similar case occurs when an instance has a habitual aspect and is produced by a third generation bilingual as Node 13 shows. However, these instances do not occur as much as the instances presented in Node 14. An example of both instances is shown in (26) and (27), respectively. (26) utilizes the verb *estar* with an imperfective non habitual aspectual feature as the speaker is reporting an unfortunate event of the speaker's uncle who lost a child at a tender age. While it is clear that this instance is assigned an imperfective aspect, it is not a habitual one. On the other hand

(27) is an example of the Imperfect used with a habitual aspect as the speaker tells a story of where he grew up and how there was a train that would pass close by. As the conditional tree shows, both of these types of examples were favored by the third generation, respectively.

(26) Tiene 23 años es como un hermano /Mhm/ por mi pero, y mi hermano tenía un hijo pero se murió cuando **estaba** chiquito.

(CESA076_324)

(27) La mayoría de los mexicanos vivían al otro lado del--de los tracks, eh, dónde **pasaba** el tren.

(CEPAVAE10M3BE)

SHL instructors should take grammar lessons from the classroom as supplementary tools rather than norms to be applied in the classroom. While grammar points and examples that textbooks demonstrate are common uses and found in different varieties of Spanish, many non-canonical uses are sometimes still presented as ungrammatical. As Beaudrie, et al. (2014) recommended, instructors should first establish what students already know (pp. 162). After having this information, instructors should focus on presenting examples that highlight the complexity of variation by providing summaries of variationist studies that present linguistic variation and showing the students' authentic data.

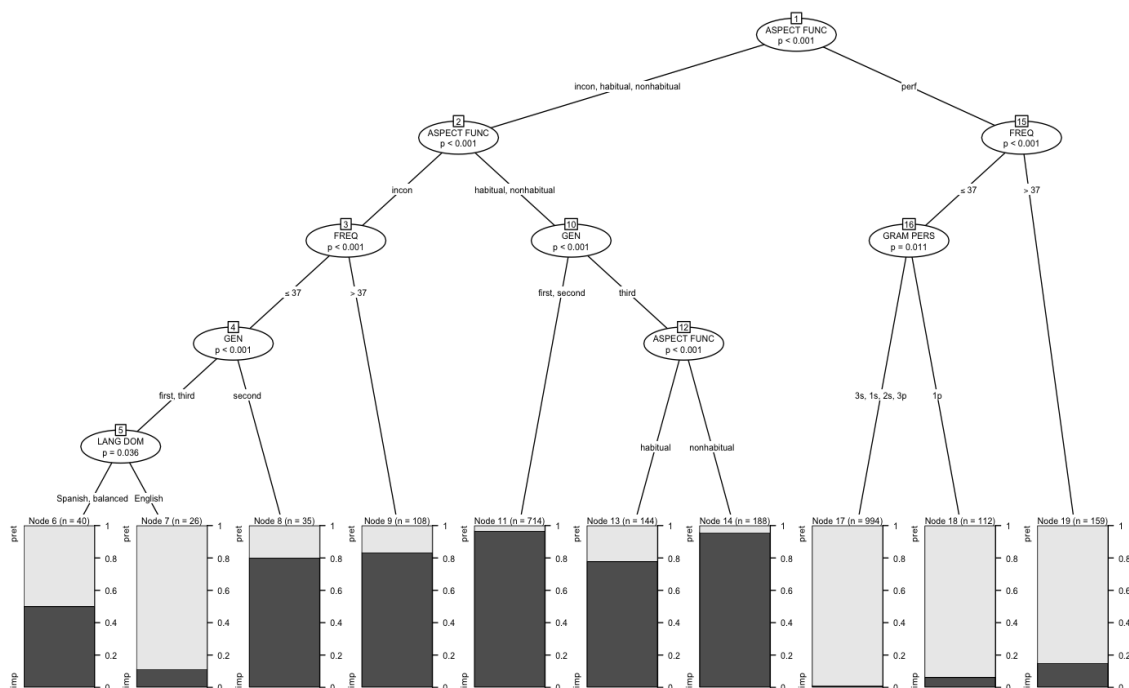


Figure 4. Conditional tree of the Imperfect

The first and second generation of speakers also utilized the Imperfect with habitual and non-habitual imperfective aspectual features. However, neither generation used either form considerably more than the other. The inconclusive instances also behaved differently than the habitual and non-habitual aspectual functions. Within the inconclusive instances, the Imperfect was preferred with verbs that were of a frequency of 37 or higher. The inconclusive instances using the Imperfect were a lot more common than inconclusive instances using the Preterite. This was a pattern that, although presented by all the generations, was more popular with the second generation. While these instances are not explored in detail in this dissertation, they deserve some attention.

An example that demonstrates what Node 8 is showing is presented in (28). The use of the Imperfect with verbs that were of a frequency of 37 or less was more likely to occur within the second generation. This example shows the speaker using the verb *hacer* which has a frequency of 19 and as we can observe in this instance depending on what exactly the speaker wanted to indicate could express an imperfective aspect as well as a perfective aspect. Instances like this ask for a further examination to be able to understand why the use of the Imperfect was preferred over the Preterite.

(28) Sí ósea, pues se- a nosotros se nos **hacía** raro porque se casaron el día de los muertos que- el día de los muertos, no para casarte. Y ya nos-nos contaron porqué nosotros “ah: está raro pero bueno.

(CESA043_1752)

The first and third generations, when influenced by language dominance, also have preferences that vary. For example, the Imperfect was preferred more by Spanish dominant and balanced bilinguals if they were part of the first and third generation and the verb was also of a frequency of 37 or less. An example of this instance is shown in (29). (29) is an example of a first generation speaker using the Imperfect with the verb *llegar* which has a verb frequency of 10. However, when generation is not influencing verb frequency, the use of the Imperfect decreases. An example of this is shown in (30) where a second generation speaker using the Imperfect with the verb *vivir* which has a lexical frequency of 18.

(29) Haz de cuenta, después de escuela ir a jugar fútbol, **llegábamos** a la casa, terminábamos la tarea y otra vez volver a hacer lo mismo. Levantarte en la mañana, ir a la escuela, salir, ir a practica y cosas así.

(CESA021_985)

(30) Ah pues es lo mismo porque ya **vivíamos** juntos así que lo única cosa que cambió era como tax information.

(CEPANS_63)

Perhaps another noticeable trend to highlight from the conditional tree, is that there are in addition some instances of the Imperfect which are utilized with a perfective aspect, especially if the verb had a frequency higher than 37. A clear example of this type of instance is shown in (31). The speaker is telling us about her husband's family, where they lived and some of the moving, they had done in the past. The speaker says that his parents had him and then they moved or came back to the city of Tucson. The context of the sentence makes it clear that the aspectual feature connected to this Imperfect verb is perfective as the mother giving birth to the child was completed. While generation was not a factor that the conditional tree presents when it comes to the Imperfect utilized with a Perfective aspect, most of the examples where this occurred came from second and third generation speakers.

(31) Emm, de su papa no. De su papa es de aquí yo creo um pero la mama vivió allá en XY y s- y y él no más- yo creo que estaban aquí los juntos, pero se fueron a

XY. **Tenían** a mi novio, o pues ahora esposo [jaja] y y um (...) y ya pues otra vez vinieron a Tucson. Y se separaron y ella se fue a México con él y luego se volvieron acá a Tucson.

(CESA024_1657)

Statistical modeling

After establishing the overall distribution of the Preterite and the Imperfect and the notable patterns as they interact with the internal predictors, the next step was to determine what predictors favor the Preterite and the Imperfect. All tokens and predictors were included in a multivariate analysis using Rbrul (Johnson, 2019). The participants and the verb lemma predictor were included as random predictors. As Table 4 shows, the multivariate analysis identified aspectual function, generation, and frequency of the verb as the significant predictors. Verb lemma was also included in the table to observe the verbs with the highest predictor weights. Only the verbs that had a factor weight of .60 or higher were included in the table.

While the predictor trees provided a visualization of the verbs that influence the use of both the Preterite and the Imperfect, Table 4 gives us exact p-values for the significant predictors. As expected, aspectual function, more specifically the perfective aspectual function is highly significant with the Preterite (p value = ~ 0). Generation is the second significant factor with a p value = .0051. The first and third generation were significant variables, with the third generation being more significant. As the frequency of the verb was input as a numeric predictor, a subsequent analysis of this verb will be

shown in the next section. However, the analysis also determined this predictor as significant with a p value = .00488.

Table 4. Multivariate analysis in Rbrul

Predictors	Tokens	Predictor weight	p-value	Range
Aspectual Function			~0	93
Perfective	1265	.99		
Inconclusive	209	.57		
Habitual	418	.12		
NonHabitual	628	.06		
Generation			.000519	37
Third	806	.67		
First	721	.53		
Second	993	.30		
Frequency of the verb +1			.00488	
Verb Lemma				
(random)	72	.85		
<i>Empezar</i>	15	.80		
<i>Pensar</i>	7	.69		
<i>Terminar</i>	29	.68		
<i>Conocer</i>	139	.66		
<i>Ir</i>	1	.64		
<i>Atender</i>	59	.63		

<i>Vivir</i>	2	.63		
<i>Batallar</i>	3	.63		
<i>Considerar</i>	4	.63		
<i>Traer</i>	1	.61		
<i>Festejar</i>	2	.60		
<i>Reír</i>				

A further look at the significant factors will be given in the next section where the data will answer the second research question that asks about the differences in the use of the Preterite and Imperfect based on generation. Since generation turned out to be a significant factor in the next section, I will provide examples of how the use of both the Preterite and Imperfect have similarities and differences demonstrated in the speech of the three different generations.

Preterite and Imperfect uses by generation

The second research question asked “Does the speaker generation influence the distributional patterns found in the first research question? To begin to answer this question, Table 5 shows the overall frequencies of the Preterite and Imperfect grouped by the three generations being examined. The data show that all generations presented a slightly higher number of tokens in the Preterite. The second generation is the generation where the Preterite and Imperfect use is more balanced. However, none of the generations had an extremely high preference for one over the other. In the next section

of the results and discussion, I present how the significant factors interact with each other.

Table 5. Distribution of Preterite and Imperfect among generations

	Imperfect		Preterite		Total
	%	N	%	N	
G1	44.4%	320	55.6%	401	721
G2	49.2%	489	50.8%	504	993
G3	44.3%	357	55.7%	449	806

Generation and aspectual function

Generation is one of the significant predictors for the Preterite and Imperfect variables. The use of the Preterite with the aspectual functions is one of the first interactions this section will highlight. Figure 5 visualizes how all three generations behave when utilizing the Preterite versus the Imperfect with the different aspectual functions. As we can observe, the first and second generation have more occurrences of the Preterite with the indeterminate aspectual tokens., whereas the second generation uses more of these indeterminate instances with the Imperfect. When it comes to the perfective aspectual function, all three generations behave similarly in that the perfective aspectual instances are preferred with the Preterite. Similarly, the non-habitual imperfective functions were mostly used with the Imperfect by all generations, and despite some variation in the first and third generation, their use is still preferred by all generations.

Perhaps the most interesting noticeable trend across generations occurs with the habitual aspectual function. As Figure 5 shows, the first generation mainly prefers to use the habitual function with the Imperfect. Similarly, the second generation speakers prefer to use the habitual aspectual function with the Imperfect, although we start seeing some variation here. Furthermore, the third generation participants demonstrate the habitual aspectual function presenting a higher variance. This means that while the habitual instances are still being presented in the Imperfect, more instances of the habitual aspect with the Preterite are appearing.

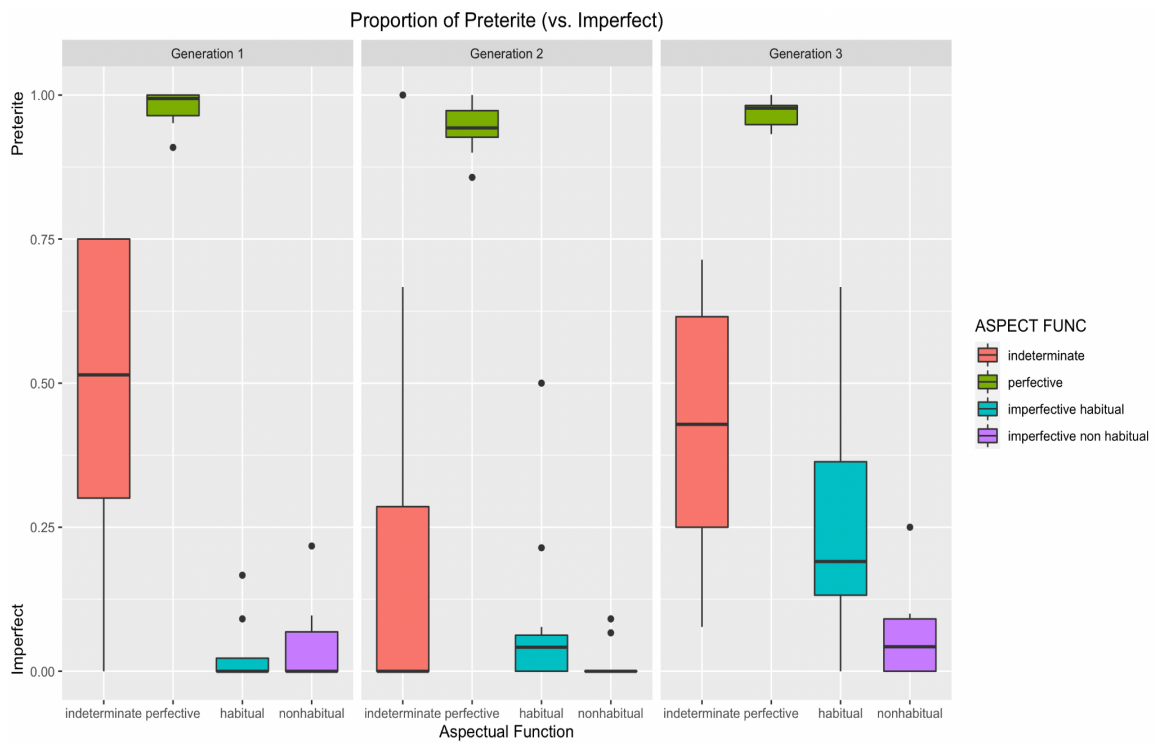


Figure 5. Proportion of the Preterite vs the Imperfect with respective aspectual categories

A mixed effects logistic regression model is shown in Table 6. This regression was completed in RStudio and searched for significance between the dependent variable, grammatical aspect, and generation X aspectual function. This regression also included the random predictors participant and verb lemma. The logistic regression shows that the interaction between second generation and habitual aspectual function as well as third generation and habitual aspectual function are statistically significant.

An example of a second generation speaker using the Preterite with a habitual aspectual function is shown in (32). This participant was describing when she would visit her friend's house, and she noticed some differences; the one she mentions in this specific example is that in her household they would always eat dinner relatively late compared to her friend's house. In this example the speaker mentions an action of eating which would happen for a period of time in the past and therefore evokes an imperfective habitual aspectual function. Similarly, we can observe a third generation speaker use the Preterite with a habitual aspectual function in (33). In this example, the speaker is remembering when he was in school and in his opinion didn't yet speak English well. Given the context this was not a one-time event nor an event where he specifically mentions it had an end as he is telling the story. The participant shares that he would be made fun of and that the other kids would laugh at him. Both examples show how this use of the Preterite with an imperfective aspectual function, more specifically a habitual aspectual function, is present and significant in these two generations.

Table 6. Fixed effects of the mixed-effects model between generation X aspectual function

	Estimate	Std. Error	Z-value	p-value
(intercept)	0.68003	0.41756	1.629	0.10340
Second generation	-1.97749	0.61132	-3.235	0.00122 **
Third generation	0.02163	0.50783	0.043	0.96603
Perfective aspectual function	3.91086	0.51847	7.543	4.59e-14 ***
Habitual aspectual function	-4.37598	0.75983	-5.759	8.45e-09 ***
Non-habitual aspectual function	-3.34968	0.53780	-6.228	4.71e-10 ***
Second generation X perfective	1.10491	0.71846	1.538	0.12408
Third generation X perfective	-0.34747	0.67298	-0.516	0.60564
Second generation X habitual	2.54589	0.95125	2.676	0.00744 **
Third generation X habitual	2.55511	0.82896	3.082	0.00205 **
Second generation X non-habitual	0.27155	1.06191	0.256	0.79817
Third generation X non-habitual	0.33952	0.72259	0.470	0.63845

(32) ... me dejaron ir a su casa era muy diferente de tipo de casa de lo que yo tenía, como nosotros como mexicanos siempre **comimos** la cena muy tarde como a las 7 o a las 8.

(CEPANS_24)

(33) Me **hicieron** mucho:, ¿cómo se dice? broma... porque no platicaba inglés bien, y todos se: **rieron**, laughing? de mí, porque no platicaba inglés muy bien, y, y pues-aquí ando aprendiendo mis primeras palabras en inglés.

(CESA006_1425)

Generation and verb frequency

Out of the 12 verbs that had a factor weight of .60 or more in the multivariate regression, more than 10 of the verbs were utilized by the first generation. These verbs are shown in Table 7 with their frequency score and proportion score based on the first generation's use. The proportion values of the verbs used in the Preterite were gathered by analyzing how many of the verbs were used in the Preterite/Imperfect. This was conducted in order to observe if the proportion of use of a verb in the first generation was low, the proportion of use in the third generation could be expected to be as low or lower due to the amount of input these other generation may have received of the verbs in the Preterite/Imperfect. In addition, those with high Preterite use in the first generation would be expected to have even higher Preterite use in subsequent generations. Out of these 10 verbs, 9 of them are highlighted due to them having a proportion of use in the Preterite over the Imperfect of .6 or higher meaning they were more likely to be used with the Preterite in the first generation. *Reír* was the only verb that in the first generation was used exclusively with the Imperfect.

Table 7. Verbs with high predictor weight with frequency and proportion values

Verb Lemma (random)	Frequency of Verb in G1	Proportion of Pret in G1
Empezar	37	.95
Pensar	3	.67
Terminar	5	.8
Conocer	5	.8
Ir	21	.67
Vivir	18	.67
Batallar	2	1
Considerar	3	1
Festejar	1	1
Reír	1	0

Out of the ten verbs provided in Table 7, seven of these verbs were used in the Preterite with a habitual aspectual feature by speakers of the second and third generation. These verbs are included in Table 8. A full table of all the verbs utilized by the first generation and their frequency and proportion is provided in Appendix C. The table provided in Appendix C shows that the highest frequency value for verb frequency in G1 is 117 and the lowest is 0. Table 8 tells us that the verbs that appeared to be significant when using the Preterite are not the verbs with the highest frequency value (37-1). This is not surprising as it was mentioned in 4.1 the most common verbs such as *ser*, *tener* and *estar* were used a lot more with the Imperfect. However, Table 8 does support the

hypothesis that many of the verbs used by the second and third generation with the Preterite and with a habitual aspectual feature, would be verbs that were already used in high proportions within the first generation. This suggests that the uses found in the second and third generation are not due to contact-induced variation. The proportion column shows that all but one of these verbs (*reír*) were highly used with the Preterite by the first generation.

Table 8. Verbs used in Preterite with habitual aspect by G2 and G3

Verb Lemma (random)	Frequency of Verb in G1	Proportion of Pret in G1
Empezar	37	.95
Pensar	3	.67
Terminar	5	.8
Conocer	5	.8
Ir	21	.67
Vivir	18	.67
Reír	1	0

This subsection demonstrated not only the distribution of Preterite and Imperfect use by generation but also provided statistical data and visualizations to help understand how these uses share some similarities and differences. While we were able to see that the second and third generation make more use of the Preterite with a habitual aspectual feature, it is important to note that these instances are not exclusive to only those two

generations. (34) shows an example of a first generation speaker exhibiting the same pattern which can suggest that this may be a change occurring in all generation but faster in the second and third generation. However, further research is needed in different Spanish speaking communities. This could include observing the behavior of these variables in different generations as well as in monolingual communities.

(34) Pues el impacto más grande mi hijita, dejar mi país y venir a Estados Unidos donde no **hablaron** el idioma (risas) que mayor impacto quiero”

(CESA074_454)

The Preterite and Imperfect: Past and Present

So far, this chapter has provided the reader with information on the distribution of the Preterite and Imperfect in the data of three generation of English Spanish bilinguals in Arizona. In addition, it has provided statistical analyses that showed the significant predictors for the Preterite and the Imperfect as well as the significant interactions highlights that the data presented. In this next section, I will answer the third research question which is “How do these results resemble previous studies conducted with monolingual and bilingual speakers?” As mentioned in Chapter 2, while there are many studies that have observed the Preterite and the Imperfect as well as other past forms, most of these studies utilize experimental methodologies instead of corpus data (especially the ones that analyze bilingual speakers). Nonetheless, I will describe similarities and differences found in the current study and previous studies and will divide them by the studies that have analyzed monolingual speech followed by the

studies that have observed these past forms in bilingual speech.

Current study and studies on monolingual speakers

The study of the Preterite and Imperfect in monolingual Spanish speakers has shown that while the use reflects patterns that are canonical and established by traditional Spanish grammars, there is also a degree of variation that exists in different varieties that are coming to light due to research. The current data show that Spanish English bilinguals in Arizona also prefer to use the Preterite with foreground information and the Imperfect with background information, similar to what previous studies have shown (Delgado-Díaz, 2014, 2018; Schwenter & Torres-Cacoulllos, 2008; Silva-Corvalán's, 1983).

Although this predictor did not turn out to be significant, the data showed that, overall, there were more instances of the Preterite than of the Imperfect. Some connections that can be made with these results are that when observing the Preterite versus the Present Perfect, the Latin American varieties (Peruvian and Mexican) highly favored the Preterite over the Present Perfect (Howe & Schwenter, 2008). All the participants in the current study had a Mexican background in that either themselves or their parents/grandparents were born in Mexico. In the present study, there were more tokens of the Preterite than of the Imperfect. However, to suggest that the Mexican varieties may prefer the Preterite over the Imperfect would need further research.

One of the findings in Delgado-Díaz (2018) was that the Preterite favored singular subject referents, while the Imperfect favored plural referents. While grammatical person as a predictor was not statistically significant, the overall distribution shows some similar trends. This study's data show that there were more instances of the Preterite with first-

person singular and third-person singular, which follows Delgado-Díaz' (2018) findings. Similarly, there were more instances of the third-person plural in the Imperfect. However, the interesting instances of the second person singular with an indefinite person interpretation mainly used in the Imperfect (73.9%) had not been noted before and did not follow known patterns. The uses of the Preterite and Imperfect and their semantic interpretations should be further studied in order to observe when speakers not only use one form over the other when referring to the second-person but also to try to understand what may prompt the speaker to utilize the Imperfect with an indefinite second-person singular pronoun. Perhaps, a future study could analyze the preference of the second-person singular versus third-person singular indefinite uses in the past. In addition, the third-person plural instances were slightly more used with the Preterite (50.5 %) rather than with the Imperfect (49.5 %). However, this difference is very small and may be due to the amount of tokens gathered.

Previous studies have also shown instances of the Imperfect being used with features that are canonically connected to the Preterite such as telic verbs, more specifically, verbs of accomplishment being used with the Imperfect (Delgado-Díaz, 2014, 2018). The data provided by the Spanish English bilinguals in Arizona did demonstrate instances of the Imperfect being used with a perfective aspectual feature⁸. However, it is important to bring the attention back to the example shown earlier in 4.1.3

⁸ Jara Yupanqui (2021) finds that the use of the Imperfect with perfective features also occurs in monolingual speakers of Amazonian peruvian Spanish. She finds that this use appears to highlight the subjective perspective of the speaker, either expressing distancing from what is narrated (epistemic use) or emotion at peak moments of the narrative plot (non-epistemic use) (pp. 87)

as seen in (35). (35) shows the verb *tener* with the Imperfect grammatical aspect and a perfective aspectual function. Typically, the verb *tener* is considered a stative verb, however, in this instance this verb could be replaced by the verb *concebir* as well as other ways to describe what the speaker is telling in her story which is that her husband's parents gave birth to her husband and then moved. Although on the surface it may appear to be a telic verb, the semantics behind it imply a level of telicity. When analyzing the speech of bilinguals, these types of examples will require a deeper analysis, which may need to be supported by lexical frequency in both languages, English and Spanish.

(35) Emm, de su papá no. De su papá es de aquí yo creo um pero la mama vivió allá en XY y s- y y él no más- yo creo que estaban aquí los juntos, pero se fueron a XY. **Tenían** a mi novio, o pues ahora esposo [jaja] y y um (...) y ya pues otra vez vinieron a Tucson. Y se separaron y ella se fue a México con él y luego se volvieron acá a Tucson.

(CESA024_1657)

Current study and studies on Spanish English bilingual speakers

The current study's participants are all Spanish English bilinguals from three different generations. Within this group of participants, we can find heritage speakers of Spanish, as well as participants who became bilingual later on in their lives. Given that this dissertation aims to avoid the idea of "correct" versus "wrong" usage, it is difficult to make connections between the results seen in the current research and other research that has analyzed bilinguals with experimental data. However, one notable similarity between

Cuza, Pérez-Tattam, Barajas, Miller and Sadowski (2013) and the current study is that the speakers (in their case, children) used the Preterite at a higher frequency than the Imperfect, similarly to what the data presented in this dissertation show. While the authors consider this an “overuse” of the Preterite and mention that it could be a transfer of the aspectual properties in English, in this dissertation I showed that many of the instances uttered by the second and third generation of the Preterite produced with an imperfective aspectual feature were instances where the verbs had a medium frequency and a high proportion of use over the Imperfect in the first generation. This could suggest that the speakers might have naturally received more input of those verbs in the Preterite and possibly less input of the Imperfect forms of these verbs; therefore, preferring that grammatical aspect over the Imperfect. Aside from making generation a predictor, language dominance was another external predictor included. This predictor did not turn out to be statistically significant, similarly to what Cuza and Miller (2015) found.

Silva-Corvalán’s (2014) study of bilingual children noted that while some of the children used the Imperfect in all contexts where the aspectual feature was imperfective, there were also some children that did not always use the Imperfect in contexts that had an imperfective aspectual function. Furthermore, the analysis showed that only one child used the Preterite in all the perfective contexts, while the rest of the children favored the Imperfect in some perfective contexts. Both of these uses are observed in the data of this dissertation. However, the variation occurring with the Preterite was more common than the variation appearing with the Imperfect. As mentioned in Chapter 2, Silva-Corvalán’s (1994) and Zentella’s (1997) seminal books were an inspiration to this dissertation. Their qualitative analyses show that variation in Los Angeles and New York exist when it

comes to using the Preterite and the Imperfect in the Spanish in the US. Both studies highlight the uses of the Imperfect with perfective aspectual features. Both authors also mention that the verbs used with these types of instances were mainly stative verbs, which is supported/corroborated by the data used in this dissertation. However, there were also many instances of stative verbs that due to the ambiguity of the aspectual function given, were considered into the indeterminate variable. These verbs will be examined in the future with other predictors.

The data analyzed in this dissertation corroborate the findings of Silva-Corvalán (1994) and Zentella (1997) by showing that the variations they demonstrate in their work are also found in the state of Arizona. In addition, the results found in the speech of Arizona speakers adds and highlights the variation that is occurring with the Preterite rather than with the Imperfect. The previous studies mentioned show an “extension” of the Imperfect, especially in situations of language contact but not much has been said about the Preterite use in these same situations up until this dissertation. One of the questions that remains is, can we find similar patterns in other varieties of Spanish in the US or in other varieties of contact?

Spanish in the US

The last research question asked, “What do these results tell us about this particular variety of Spanish in the U.S?”. The present dissertation analyzed the uses of the Preterite and Imperfect in the variety spoken by English bilinguals in the state of Arizona. As mentioned in Chapter 3, these speakers all had lived in Arizona for a minimum of five years. The Southwest of the US has been an area where sociolinguists

have conducted plenty of research on linguistic characteristics, many times with the goal of understanding whether the linguistic features can be attributed to language contact or to internal changes. Some of the most studied linguistic variables in the US are copula choice, subject pronoun, expression, progressive constructions, and mood choice (Bessett & Carvalho, 2021). The results presented in this dissertation add to the previous research on the study of morphosyntactic features pertaining to the Spanish spoken in the Southwest of the US. Similarly to previous sociolinguistic studies of the morphosyntactic variables mentioned above, the characteristic and variation present in the Spanish spoken by the speakers in Arizona shares a progressive pattern in all three generations. While this dissertation did not compare the use of the Preterite and Imperfect to a monolingual Mexican variety, also known as a reference lect (Otheguy & Zentella, 2012), the aim was to be able to observe similarities and differences occurred between generations. The results show a systematic pattern where all generations utilize the Preterite and the Imperfect in prototypical or canonical ways. However, the data also show that these Spanish speakers who live in the US also make use of the Preterite with imperfective aspectual features., as well as some uses of the Imperfect with a perfective aspectual feature. While this is heightened in the second and third generation, it is important to note that it also occurs in the first generation. Therefore, there is systematicity found in their use of both forms, hardly promoted by influences external to Spanish.

The results provided by this dissertation may suggest that there are other varieties that are evolving and making aspectual choices that are different than the canonical forms typically taught in the Spanish classroom. While there is no need to compare to monolingual varieties in order to understand the complexity and systematicity of the

Spanish spoken in the US, this study's results may help researchers question if variation is present in other Spanish varieties. As mentioned in section 4.3 even in monolingual context instances of the Imperfect have been noted with a perfective aspectual feature. The Imperfect does allow a certain higher degree of range that the Preterite may not appear to possess. However, this dissertation shows there is variation when using the Preterite. The results suggest the following: The pragmatics behind habituality found in these US Spanish speakers may vary when compared to the pragmatics behind other Spanish speech communities. While a habitual state or event occurs in the past for a period of time, perhaps the weight of the state or event being in the past and therefore completed helps these speakers make the Preterite and habitual aspect connection.

Pedagogical Implications

Understanding the variation that is present in all Spanish varieties should and must come with pedagogical implications for the Spanish classroom, especially in the case of the US. Given that Spanish speakers and heritage speakers are a minoritized community in this context, understanding the sociopolitical context that surrounds their language varieties should occur in the classroom (Holguín Mendoza, 2021; Loza & Beaudrie, 2021; Loza, 2021) At the very least, as educators and researchers, we should aim to remove societal stigmas from the classroom. In this case, I will focus on a few pedagogical implications and considerations for the Spanish as a heritage language (SHL) classroom.

Although an SHL classroom encompasses much more than just the teaching of language in a specific community, the subject of grammar always tends to be present.

According to Beaudrie, et al. (2014), the concept of grammar can have multiple interpretations. One of the interpretations of grammar refers to the mental system that the speakers of a community share. A second interpretation refers to being able to explain the uses of a grammatical terminology and is related to metalinguistic knowledge. A third interpretation emphasizes what is considered "appropriate" and refers to the "standard" usage (Beaudrie et al., 2014). Using the findings of this dissertation, instructors should emphasize the importance of the first interpretation to grammar in the classroom which focuses on discussing grammar as a non-static or generalized concept. To aid with this presentation of grammar in the classroom, instructors can discuss with students the variation that exists and that has been found in studies such as this one. This would help students keep away from notions of grammar such as the third one and instead learn a range of grammar possibilities that exist within different Spanish communities. In addition, this approximation to grammar can highlight that the varieties the SHL learners bring to the classroom are not and should not be considered incomplete.

Although linguistic variation appears to be welcomed in the classroom, much of the spotlight appears to be shed on the lexical characteristics rather than morphosyntactic characteristics. I argue that shedding light on morphosyntactic characteristics of different Spanish varieties, with a specific emphasis on US Spanish is absolutely necessary not only to bring language awareness to the students but also to try to get rid of language ideologies that make students believe their varieties are deficient. As the results of this dissertation show, grammatical forms as common as the Preterite and the Imperfect present variation in the spoken Spanish variety of speakers in the US; a type of variation that unfortunately is not presented in textbooks where much emphasis is pointed on the

‘standard’ peninsular variety (Padilla & Vana, 2019). Data such as the ones found in the present dissertation and Delgado-Díaz, (2014) can become useful tools to utilize in the classroom.

SHL instructors should take grammar lessons from the classroom as supplementary tools rather than norms to be applied in the classroom. While grammar points and examples that textbooks demonstrate are common uses and found in different varieties of Spanish, many non-canonical uses are sometimes still presented as ungrammatical. As Beaudrie, et al. (2014) recommended; instructors should first establish what students already know (pp. 162). Following their suggestion, when discussing grammar topics such as the Preterite/Imperfect, teachers begin with a discussion on what students already know of these forms. If students mention what may appear as non-canonical uses to the instructor (forms that may not be presented in a textbook), the instructor should not immediately attempt to correct the student (Loza, 2021). Instead, the instructors can begin by welcoming what the students know and continue from there. A class activity could include analyzing what the textbook mentions are the differences between the uses of the Preterite and the Imperfect and acknowledging that those uses are most likely the standardized forms that speakers are expected to use. However, a follow-up activity could have students see that the Real Academia Española (RAE) has overtime changed their definition of each past form across times. Additionally, RAE (2010) presents other uses for the Preterite and Imperfect that go beyond what the typical textbook shows. Having gone through these two activities, instructors can then present examples of the variation that exists in different Spanish varieties. For example, in spoken Puerto Rican Spanish, Delgado Díaz (2014) shows that

the Imperfect can be used with accomplishment verbs which are typically associated with the Preterite. In addition, instructors can mention that the use of the Preterite with habitual imperfective interpretations is present in speakers in the US with a higher use in the second and third generation. Table 9 shows an example of how this variation can be simply showed to the students in the classroom.

Table 9. Presenting Preterite and Imperfect variation to HL learners

	Variation within the Imperfect	Variation within the Preterite
Where	Puerto Rican Spanish (Delgado-Díaz, 2014)	US Spanish (Present study)
Examples	En el tiempo de antes, este, habían peleítas, sí, sí de nenes chiquitos, pero que a los dos o tres días se buscaban uno a los otros, pero ahora no.	Pues el impacto más grande mi hijita, dejar mi país y venir a Estados Unidos donde no hablaron el idioma, que mayor impacto quiero

Many times, students are concerned with the grammar of their variety and often think that it ‘lacks grammar’. If instructors can show examples that demonstrate US varieties have grammar just like other varieties, this can start the conversation on the complexity and systematicity of language.

Utilizing authentic data in the classroom is something that is already utilized in Spanish linguistic courses. If pedagogues give students the opportunity to observe, analyze and discuss linguistic variation past the lexical level, this could not only help students banish negative language ideologies they may have towards their varieties or

other varieties, but it may also help increase students' interest in continuing their studies in the language. Studies like this present one can bring attention to linguistic characteristics of US Spanish can be utilized in the classroom by providing authentic material for the instructor to use.

In addition, studies on US Spanish can help instructors know more about the varieties their students are bringing to the classroom and utilize this knowledge when dealing with aspects of teaching such as corrective feedback (CF). Loza (2021) argues that “oral CF should be limited to explicit Focus on Form (FonF) instruction that provides students with authentic materials and discourses relevant to their own cultural backgrounds and educational interests” (pp. 131). This type of oral CF focuses on utilizing materials and discourse that are relevant to the students' varieties and not idealized forms that do not form part of students' cultural experiences. Loza (2021) explains that adding this strategy to a CLA-based curriculum could turn oral CF from “corrections needed to speak more proper Spanish” to “instructional modifications” (pp.131). Employing this type of pedagogical intervention can make the expanding knowledge of the SHL learner more student-focused rather than purely language-focused in a prescriptive sense.

Lastly, Holguín Mendoza (2021) notes that the first step towards achieving social justice in the classroom is to “fully explore our own biases and linguistic ideologies in relation to race, ethnicity, class, gender, ability, sexuality, and indigeneity (among many others)” (pp.153). The understanding and appreciation of linguistic variation can be very difficult for many instructors if they have not explored their biases towards these aspects that are all intertwined with language. In addition, I suggest that as researchers we should

self-reflect on the goal of our research: Are we trying to bring this knowledge back to the classroom to help our students? How are our own ideologies reflected in our research? These are two questions that I argue will help promote the goal of linguistic equality and linguistic empowerment for the SHL students and speakers.

Summary

This chapter presented the results of the variationist analysis of the Preterite and Imperfect in the three generations Spanish English bilinguals in Arizona. Descriptive and inferential statistics were presented to demonstrate the trends and significant factor that favor both the Preterite and the Imperfect. In addition, examples were provided to address the data findings. The descriptive statistics showed that the usage was relatively balanced between the Preterite and the Imperfect. In addition, the data showed that all generations used canonical uses of the Preterite and Imperfect and aspectual function was the main predictor that favored both past forms.

The results show that variation was also present, especially within the second and third generation. Instances of the Preterite with a habitual aspectual function were instances that although present in all generations, increased with each subsequent generation. This habitual aspectual function used with the Preterite was significant within the second and third generation. The results suggest that the pragmatics behind habituality found in these US Spanish speakers may vary when compared to the pragmatics behind other Spanish speech communities. Even though a habitual state or event occurs in the past for an imprecise period, these speakers may be able to connect

the Preterite and habitual component since the state or event is in the past and therefore complete.

The discussion section of this chapter focused on interpreting how the results provided by this dissertation make a significant contribution to the study of US Spanish and its speakers. In addition, the chapter provided pedagogical implications for the SHL classroom as well as suggestions for researchers and pedagogues that research or teach SHL speakers and learners. This section of the chapter had the goal to help instructors bring critical language awareness to themselves and the students.

CHAPTER 5

CONCLUSION

The present dissertation utilized oral corpus data to conduct a variationist study on the use of the Preterite and Imperfect in Spanish English bilinguals living in central and southern Arizona. As the study of variable past forms is generally scarce from a variationist and sociolinguistic perspective in both monolingual and bilingual varieties, this is the first variationist study to observe the alternation of both past forms in the state of Arizona. This dissertation first provided background on what is known about the Preterite and the Imperfect. Then it explained the justification and motivation for the present study by showing previous literature and gaps related to the study of the Preterite and the Imperfect, as well as in the studies of Spanish in the US. In addition, I suggested some pedagogical implications to bring to the heritage language classroom where US Spanish is or should be a major topic of conversation.

This last chapter begins by providing a summary of the dissertation as a whole along with the main findings. Second, the chapter highlights the main contributions this dissertation makes to the field of variationist research as well as the study of US Spanish. Third, I describe the limitations of the current study as well as considerations for the envelope of variation. Lastly, I provide ideas for future directions on studies of the Spanish past tense and the verbal system of US Spanish.

Summary

This dissertation aimed to answer the following research questions:

- A. What is the distribution between the Preterite and the Imperfect in the Spanish of the Spanish and English bilinguals in Arizona? What are linguistic and extra-linguistic factors that predict the use of the Preterite and Imperfect?
- B. What similarities and differences are present in the use of the Preterite and Imperfect based on the generation of the speaker?
- C. How do these results resemble previous studies conducted with monolingual and bilingual speakers?
- D. What do these results tell us about this variety of Spanish in the U.S?

To address these questions a sociolinguistic study was conducted with the interviews of 30 English Spanish bilingual speakers living in Arizona. The 30 sociolinguistic interviews were gathered from CESA (Carvalho, 2012-) and CEPA (Cerrón-Palomino, 2012-) corpus. Instances of the Preterite and the Imperfect were coded to analyze their distribution and the linguistic and extra linguistic predictors. The linguistic predictors analyzed were *aspectual interpretation, clause type, grammatical person and number, verb lemma and verb frequency*. The extra linguistic factors analyzed were *generation, language dominance, age and sex*.

The distribution of the Preterite and Imperfect in the data revealed that both forms were used a similar number of times with the Preterite (53.7%) being used slightly more than the Imperfect (46.3%). The results of the quantitative analysis revealed that in order

of magnitude, aspectual function, generation, and frequency of the verb were the factors that predicted the Preterite and the Imperfect. While most Preterite and Imperfect uses followed the known prototypical uses of these forms, an interaction between generation and aspectual function showed significance when the Preterite is used with a habitual aspectual function by both the second and third generation. This dissertation was able to show that the Preterite and Imperfect carry a degree of variation that goes beyond the traditional understandings of these forms, and is similar to what the few other studies analyzing the Preterite, and Imperfect have shown (e.g., Silva-Corvalán, 1994; Delgado-Díaz, 2014, 2018; Zentella, 1997)

Contributions to the field of language variation research

To date, very few studies have examined the use of Preterite and Imperfect using a variationist approach (Delgado-Díaz, 2014, 2018). As the few previous studies on the Preterite and the Imperfect have suggested, the present dissertation shows that the Preterite and the Imperfect are not as stable as previously stated by traditional grammar and current language textbooks. As the current and previous studies that have observed the Preterite and the Imperfect from a sociolinguistic perspective have noted, variation is present in all of these varieties of Spanish. Therefore, acknowledging and bringing attention to this linguistic variation is important and can aid Spanish English the existing prejudice around different varieties of US Spanish.

This dissertation has highlighted the relevance and importance of variationist studies and the field of language variation. This research paradigm has shown that variation is inherent to language, and is present not only in every language, but also in

every variety of a language, a key finding in rejecting the idea of linguistic homogeneity and “correctness” that traditional grammars tend to portray. While this dissertation only focused on the dichotomous pair of the Preterite and the Imperfect, it is a steppingstone for future research on the variation that could be found among other past forms.

Following a variationist approach, this study analyzed linguistic and extra-linguistic factors. Given the scarcity of variationist studies on the Preterite and the Imperfect, a selection between factors groups that have appeared to be relevant in previous variationist studies as well as other predictors were examined. In the case of this dissertation, the linguistic predictor that was most salient was the aspectual feature. The extra-linguistic factor group that stood out was generation and the random factor group that also was highlighted in the data was verb lemma, which factor was also observed alongside the verb frequency predictor. However, the limitation section will address the predictors that were not observed in this dissertation and may need to be observed in the future.

In terms of data analysis, this dissertation utilized a mix of Rbrul (Johnson, 2019) and Rstudio (RStudio Team, 2020) to be able to visualize and present the data. The advantages of Rbrul, which is typically utilized when analyzing comparative variationist data, is that it clearly presents the values that variationist studies present when interpreting the inferential statistics. However, Rstudio can more easily provide plots and graphs that can help visualize the data (Wickam, 2009). In both cases, the verb lemma and participant factor groups were input as random factors, respectively when running the linguistic and social mixed-effects regressions.

Contributions to the field of US Spanish

This is the first variationist study that explores these forms with Spanish English bilingual speakers in the US. This is extremely important in order to be able to demystify the idea that the Spanish spoken in the US, in this case, Arizona, lacks purity, complexity, or is incomplete (Bessett & Carvalho, 2021). To date, some of the most studied morphosyntactic variables are the subject pronoun expression (e.g., Bessett, 2018; Cerrón-Palomino, 2016; Hurtado, 2001; Limerick, 2017; Otheguy & Zentella, 2012; Torres Cacoullos & Travis, 2018; among others), copula choice (e.g., Bessett, 2012; Gutiérrez, 1992; Silva-Corvalán, 1994, among others), progressive constructions (e.g., Carter & Wolford, 2018; Chaston, 1991; Klein, 1980; Lamana, 2008; Marques Martinez, 2009; Wilson & Dumont, 2015; among others), and mood selection (Lynch, 1999; Martillo Viner, 2016, 2018; Silva-Corvalán, 1994; Torres, 1989; among others). This dissertation adds to this body of literature by providing further insights on the topic of past form selection in US Spanish from a variationist perspective and suggests that more research be done on the topic with different Spanish speakers in the US.

Although a comparison was not made between US Spanish and a monolingual variety as typically seen in comparative variationist studies, the results show that the distributional patterns in all three generations share some similarities as well as differences. The choice to not compare Spanish English bilingual speech to Spanish monolingual speakers was made for a couple of reasons. On one hand, this dissertation aimed to obtain a first look of how these two past forms were presented in the speech of Arizona and observe how this US variety compared to previous findings in other US Spanish communities such as California (Silva-Corvalán, 1994) and New York (Zentella,

1997). On the other hand, this was done to avoid perpetuating the fallacy that bilinguals are just two monolinguals in one (e.g. Pascual y Cabo & Rothman, 2012; Grosjean, 2006; among others). Many times, when this comparison is made, the monolingual variety is assumed as the model variety. For that reason, this dissertation focused solely on understanding how the Preterite and Imperfect are presented by different generations of US Spanish speakers in Arizona. In addition, noting that the use of the Preterite with habitual instances was significant in the third generation suggests that there may be a change in progress. However, more research is needed as well as the exploration of different linguistic and extra-linguistic factors when observing the Preterite and Imperfect.

This dissertation has also demonstrated the importance of visibility of US Spanish with the goal of spreading the notion that it should be observed in its own right just as other Spanish-speaking varieties have been studied. Additionally, as recent research on monolingual Spanish varieties has shown that variation between the Preterite and the Imperfect exists, this dissertation has also shown that variation is present in this US variety and rejects the assumption that US Spanish is a mixed variety. This was demonstrated in the data as canonical uses of both the Preterite and the Imperfect were present in all generations, by all speakers. However, similar to other varieties that have found variation, this US Spanish variety presents its own variation in the form of the Preterite being used with the imperfective habitual aspect. As Otheguy (2016) has argued, US Spanish should not be viewed or referred to as having incomplete grammar and makes the call for viewing dialectal differences in US Spanish as the indication of intergenerational linguistic changes (Bessett & Carvalho, 2021; Otheguy, 2016).

The US Southwest, the area with the closest proximity to Mexico, has been one of the more prominent and popular areas for studies of US Spanish because of the intense language contact occasioned by northbound migration from the south . However, it is important to continue the research on different linguistic variables in different areas of the country as this can also give us an insight into the evolution that may be occurring in US Spanish. This is important to determine whether the proximity and frequency of interaction with nearby family members and communities have a particular impact on the characteristics found in the US Southwest. Therefore, within the US, the degree of exposure to Spanish monolinguals can be used to provide more insights on the linguistic continuities found in the Spanish speakers of the Southwest (Bessett, 2012, 2015, 2018; Bessett & Carvalho, 2021; Cerrón- Palomino, 2016).

Contributions to the field of Spanish as a heritage language

This dissertation has contributed to the field of Spanish as a heritage language by providing pedagogical implications that can be drawn from variationist studies. Additionally, as mentioned in the previous section it has provided insights on the use of the Preterite and Imperfect by bilingual speakers. Within these bilingual speakers, heritage speakers of Spanish can be found in the second and third generation. While a distinction outside of generation and language dominance was not provided, the characteristics as well corpus data included as examples in this dissertation can help instructors reflect on the complexity of the language variation present in heritage varieties. In addition, the examples and findings in this dissertation can be utilized in Spanish heritage courses when the topic of the past tense is mentioned in the classroom.

These types of examples can help highlight the importance of how language is spoken rather than how it should be spoken.

The unification of variationist studies and heritage language pedagogy may be one important task that can help students and even instructors get rid of different negative language ideologies that are unfortunately commonly present in the heritage language classroom. I suggest that the more instructors can teach their students about findings that variationist studies such as this dissertation provide, the more the idea that certain varieties of Spanish not being as “pure” or “correct” can be eradicated. While this is not a suggestion that will solve the complex issues around language ideologies, it can be beneficial to dismantle certain ideologies such as the standard language ideology. Doing this would be a great service given that the standard language ideology is one of the most prominent language ideologies seen among heritage language students and their instructors.

Aside from highlighting the importance of utilizing authentic data that is relevant and of close connection to heritage students, the pedagogical implications section (see Chapter 4) proposed how the findings of this dissertation can be implemented in the use of oral corrective feedback (see Loza, 2021). This recommendation continues the conversation on the importance of making the heritage language classroom a place where learning can be centered around how language is spoken by specific communities and avoiding the idea that language should be spoken a specific way. In addition, I suggested that following a Critical Language Awareness pedagogy should be useful to highlight the complexity of variation that exists and help students develop their critical language awareness, one of the goals of HL instruction (see Beaudrie & Wilson, 2022).

Limitations

This dissertation is not free from limitations. The main drawback is the number of participants. The participants in the data represent three generations of English Spanish bilingual speakers. In order to have a balanced dataset with the same number of participants per generation, a search was done through the corpora to find participants that fit the generation descriptions. While it was useful to combine the CESA and CEPA corpora which provided a pool of both Central and Southern Arizona speakers, there were less speakers of the third generation in both corpora. Not surprisingly, it is overall harder to find third generation Spanish speakers. This can of course be attributed to the fact that language shift takes place even in states with high Hispanic populations (American Community Survey, 2019). The number of speakers was also balanced by the sex of the speaker which made choosing the participants out of the corpora more difficult. However, while it would be beneficial to be able to have more third generation speakers, a larger overall sample size would be needed. Given that the speakers in the Southwest share some similarities (in contrast to New Mexican speakers) it may be useful to create a larger sample size that can include other corpora from the Southwest.

Variationist analyses that include the Preterite, and the Imperfect are still scarce and not much is known about either of these forms and how they are actually used by both monolinguals and bilinguals. While research on verb aspect variationist still at its inception unlike, for instance, subject pronoun expression, the study of solely the Preterite versus Imperfect dichotomous pair is another one of the limitations. By solely analyzing this dichotomous pair, only a certain part of the past tense is being explored and this leaves another section of the Spanish past system that is not being taken into

consideration. This dissertation sought to create a steppingstone for the investigation of the past in US Spanish in order to provide an account of the Preterite and Imperfect uses and trends. However, a consideration of all or more aspectual forms of the past is needed to achieve a more complete understanding of the past tense and its use.

Another limitation of this dissertation is the lack of certain predictors that have been found to be significant as well as predictors that are more closely related to pragmatics. Vendler's aktionsart was one of the predictors that was not analyzed in this dissertation and instead the categories of aspectual feature were observed. While Vendler's verb classifications have had some criticism (De Miguel, 1999, Dowty, 1972, Verkuyl, 1972) they could be useful to examine/study/investigate in the future. In addition, while some researchers have added cognitive and perception verbs to Vendler's verb classification (Aaron, 2006, De Miguel, 1999), the classification of lexical semantics might become useful and perhaps the more detailed the classification the better the analysis would turn out. Another predictor that was not coded was temporal adverbs. This predictor has typically analyzed durative, punctual, and iterative verbs as well as the presence of no adverb (Delgado-Díaz, 2014, 2018). However, when coding, the data showed that when there was no adverb, there could still be other information surrounding the Preterite or the Imperfect conveying a sense of either time or space. Given this complexity, the researcher chose to explore these connections in a future study.

Future Directions

Some future directions have already been mentioned in the limitations section. However, I will provide further future directions based on what was presented in this

dissertation. The main future direction to expand this line of research is to examine other aspectual forms of the past in a similar community in the Southwest in order to get a more complete analysis of the use of the Spanish past. A look at all these forms in other highly concentrated Hispanic states such as Florida or New York would be helpful to be able to learn the similarities and differences that may be present among other US Spanish speakers. Similarly, a look at the past forms in bilingual speakers that have a heritage other than Mexican would be beneficial. Perhaps a good start would be focusing on Puerto Rican English Spanish bilinguals on the East Coast. This type of research could be compared to what Delgado-Díaz (2014) found in his Puerto Rican monolingual speakers.

In terms of future directions on US Spanish, there are many linguistic forms that have not been explored from a variationist perspective or even from a sociolinguistic perspective as most of the studies that include US Spanish speakers utilize experimental methodologies. A necessary future direction for the field of US Spanish to continue to grow would be to continue to search for possible linguistic variables that have been analyzed in monolingual settings but not in the Spanish English bilingual setting in the US. One example of this would be the use of the Present Perfect in US Spanish. As a linguistic variant there is considerable knowledge about the use of the Present Perfect in Peninsular and Latin American Spanish, and it is a variant that has shown interesting conditioning. This could be part of a future past form analysis in US Spanish.

Another topic of interest might be the difference between production and perception of the past tense. This could be a study where real oral data could be supplemented by an experiment. However, the real data that are provided by an interview should still be the main point of reference for the production section of the study. While

an analysis of production and perception would require a methodology with longer implementation, it could be interesting to observe if there are certain features presented in one versus the other.

A last future direction for variationists or researchers who are interested in studying US Spanish would be to encourage them to bring their research findings to the classroom. As linguists we must never forget that while we focus on the language itself, that language is produced by people to communicate, and those communicative practices should be respected. In addition, the people should be our main focus when thinking about the outcome of our research. In the case of US Spanish, researchers are dealing with a minoritized community within the US. In the case of immigrant participants, these participants have many times felt like outcasts in US society. On the other hand, in the case of the heritage speakers of Spanish, these speakers have many times felt rejected by not one but two cultures—both the US and their heritage culture. In the case of the heritage speakers of Spanish, if an instructor as a researcher has the opportunity to explore linguistic features of their own language variety, bringing this to the classroom to provide awareness can not only help students feel seen but can also help create more interest from the students towards their variety. As long as the research being conducted sheds a positive light on the complexity and beauty of language variation, students will be able to understand that variation goes beyond lexical and phonological characteristics and view their variety and other varieties more positively.

As the first variationist study of the Preterite and Imperfect used by Spanish English bilinguals, this dissertation has provided insights of the use of these past forms in bilingual speakers and contributed to the research on US Spanish. The results have shown

that these speakers from Arizona use both the Preterite and Imperfect in canonical and non-canonical forms through all three of the generations with some trends becoming clearer in the second and third generation. There is still much we don't know about US Spanish and its speakers. Questioning forms that have for long appeared to be static is important not only for the field of language variation but also for the speakers themselves who may have been corrected or judged for speaking the way they do. With this dissertation I hope I have sparked the interest on other similar forms and their use in US Spanish.

REFERENCES

- Aaron, J. 2006. Variation and change in Spanish future temporal expression: rates, constraints, and grammaticization. New Mexico, NM: The University of New Mexico dissertation.
- Alarcón, I. V. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14(3), 332-350.
- Alcina Franch, J., & Blecua, J. (1975). *Gramática española*. Ariel.
- Ayoun, & Salaberry, M. R. (2005). Tense and Aspect in Romance Languages: Theoretical and applied perspectives. In *Tense and Aspect in Romance Languages*. John Benjamins Publishing Company.
- Bardovi-Harlig, K. (2000). Tense and aspect in second language acquisition: Form, meaning, and use. *Language Learning: A Journal of Research in Language Studies*, 50, 1.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Mixed-effects models using lme4. *J Stat Softw*, 67(1), 1-48.
- Beaudrie, S. (2012). Research on university-based Spanish heritage language programs in the United States: The current state of affairs. In S. Beaudrie & M. Fairclough (Eds.), *Spanish as a heritage language in the United States: The state of the field* (pp. 203-221). Georgetown University Press.
- Beaudrie, S., Ducar, C. & Potowski, K. (2014). *Heritage language teaching: Research and practice*. McGraw Hill.
- Beaudrie, S. M., & Wilson, D. V. (2021). Reimagining the Goals of HL Pedagogy through Critical Language Awareness. In *Heritage Language Teaching* (pp. 63-79). Routledge.
- Bello, A., & Trujillo, R. (1847). *Gramática de la lengua castellana: destinada al uso de los americanos*. Instituto Universitario de Lingüística Andrés Bello.
- Bessett, R. M. (2012). *La extensión de estar a través del tiempo y el espacio: un estudio diacrónico de los siglos XIII–XX y sincrónico de Hermosillo, Sonora y el sur de Arizona*. (MA thesis). University of Arizona, Tucson.
- Bessett, R. M. (2015). The extension of *estar* across the Mexico-US border: Evidence against contact induced acceleration. *Sociolinguistic Studies*, 9 (4), 421–43.

- Bessett, R. M. (2018). Testing English influence on first person singular ‘yo’ subject pronoun expression in Sonoran Spanish. In J. E. MacDonald (Ed.), *Issues in contemporary trends in Hispanic and Lusophone linguistics* (pp. 355–72). Amsterdam: John Benjamins.
- Bessett, R. M., & Carvalho, A. M. (2021). The Structure of US Spanish. In *Heritage Language Teaching* (pp. 44-62). Routledge.
- Bybee, J. L., Perkins, R. D., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Bybee, J. (2001). Frequency effects on French liaison. *Typological studies in language*, 45, 337-360.
- Bybee, J., & Torres, R. (2008). Phonological and grammatical variation in exemplar models. *Studies in Hispanic and Lusophone Linguistics*, 1(2), 399-414.
- Canagarajah, S. (2011). Codemeshing in Academic Writing: Identifying Teachable Strategies of Translanguaging. *The Modern Language Journal*, 95(3), 401–417.
- Caravedo, R. (1989). *El español de Lima : materiales para el estudio del habla culta*.
- Carvalho, A. M. (2012-). *Corpus del Español en el Sur de Arizona (CESA)*. University of Arizona. Cesa.arizona.edu.
- Carvalho, A. M. (2016). The analysis of languages in contact: A case study through a variationist lens. *Cadernos de estudos linguísticos*, 58(3), 401-424.
- Cerrón-Palomino, Á. (2012-). *Corpus del Español de Phoenix, Arizona (CEPA)*. Arizona State University.
- Cerrón-Palomino, Á. (2016). (No) English interference on US Southwest Spanish?: A look at variable subject pronoun expression in phoenix Spanish English bilinguals. *Sociolinguistic Studies*, 10 (3), 383–408.
- Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge University Press.
- Comrie, B. (1985). *Tense*. Cambridge University Press.
- Cuza, A., Pérez-Tattam, R., Barajas, E., Miller, L., & Sadowski, C. (2013). The development of tense and aspect morphology in child and adult heritage speakers. *Innovative research and practices in second language acquisition and bilingualism*, 38, 193-220.

- Cuza, A., & Miller, L. (2015). The protracted acquisition of past tense aspectual values in child heritage Spanish. *Hispanic Linguistics at the Crossroad: Theoretical Linguistics, Language Acquisition, and Language Contact*, 211-30.
- De Garavito, J. B., & Valenzuela, E. (2006). The status of ser and estar in late and early bilingual L2 Spanish. In *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 100-109). Somerville, MA: Cascadilla Proceedings Project.
- De Miguel, Elena. 1999. El aspecto léxico. *Gramática descriptiva de la lengua española*, ed. By Ignacio Bosque and Violeta Demonte, 2977-3060. Madrid: España.
- Delgado-Díaz, G. (2014). Teoría versus uso: análisis sobre el pretérito y el imperfecto. *Boletín de filología*, 49(1), 11-36.
- Delgado-Díaz, G. (2018). Dialectal variation of the preterit and Imperfect: A closer look to Puerto Rican and Buenos Aires Spanish. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 31(1), 64-93.
- Delgado-Díaz, G. (2021). *The evolution of Spanish past forms*. Routledge.
- Díaz-Campos, M. (Ed.). (2022). *The Routledge Handbook of Variationist Approaches to Spanish*. Routledge, Taylor & Francis Group.
- Dowty, David. 1972. Temporally restrictive adjectives. *Syntax and Semantics 1*. 51-62.
- Dunn, A. and Fox Tree, J. E. (2009) A Quick, Gradient Bilingual Dominance Scale. *Bilingualism: Language and Cognition* 12(3): 273–289.
- Española, R. R. A. (2010). *Nueva gramática de la lengua española manual*. Espasa.
- García, O., & Li, W. (2014). *Translanguaging : language, bilingualism and education*. Basingstoke, Hampshire: Palgrave Macmillan.
- Geeslin, K. L., & Gudmestad, A. (2010). An exploration of the range and frequency of occurrence of forms in potentially variable structures in second-language Spanish. *Studies in Second Language Acquisition*, 32(3), 433-463.
- Gili Gaya, S. (1960). *Curso superior de sintaxis española*. Barcelona: Vox.
- Gutiérrez, M. J. (1992). The extension of *estar*: A linguistic change in progress in the Spanish of Morelia, Mexico. *Hispanic Linguistics*, 5, 109–41.

- Heine, B., & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge University Press.
- Hopper, P. J. (1979). Aspect and foregrounding in discourse. In *Discourse and syntax* (pp. 211-241). Brill.
- Mendoza, C. H. (2021). Sociolinguistic Justice and Student Agency in Language Education: Towards a Model for Critical Sociocultural Linguistics Literacy. In *Heritage Language Teaching* (pp. 138-156). Routledge.
- Howe, C., & Schwenter, S. A. (2008). Variable constraints on past reference in dialects of Spanish. In *Selected proceedings of the 4th Workshop on Spanish Sociolinguistics* (Vol. 1, pp. 100-108). Somerville, MA: Cascadilla Proceedings Project.
- Hurtado, L. M. (2001). La variable expresión del sujeto en el español de los colombianos y Colombo-Americanos residentes en el condado de Miami-Dade. (Doctoral dissertation). University of Florida, Gainesville.
- Jara Yupanqui, M. (2021). Subjetividad e Intersubjetividad de los Diminutivos en-it en el Español Amazónico Peruano. *Revista Española De Lingüística*, 51(1), 23.
- Kim, J. Y. (2012). Perception of lexical stress by L2 learners of Spanish and heritage speakers. In *Poster presented at the 4th Annual SLATE Graduate Research Symposium, University of Illinois at Urbana-Champaign, IL, USA*.
- Klein, W., & Von Stutterheim, C. (1987). Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte*, 109, 163-183.
- Lavandera, Beatriz (1978). *Lo Quebramos, But Only in Performance*. [Washington, D.C.]: Distributed by ERIC Clearinghouse.
- Labov, W. (1963). The social motivation of a sound change. *WORD*, 19, 273-309.
- Labov. (1966). *The linguistic variable as a structural unit*. ERIC Clearinghouse.
- Labov, W. (1973). *Sociolinguistic Patterns*. United Kingdom: University of Pennsylvania Press, Incorporated.
- Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer (Eds.). *Language in use: readings in sociolinguistics* (pp. 28–53). Englewood Cliffs, NJ: Prentice-Hall.
- Lope Blanch, Juan M. 1976. *El habla popular de México: Materiales para su estudio*. México: Universidad Nacional Autónoma de México (UNAM).

- Lamanna, S. (2008). Usage of Imperfect and Imperfect Progressive Verb Forms in Spanish as a Majority and Minority Language: Is There an Effect for Language. In *Selected Proceedings of the 10th Hispanic linguistics symposium* (pp. 251-264).
- Lamanna, S. G. (2012). *Colombian Spanish in North Carolina: The role of language and dialect contact in the formation of a new variety of US Spanish* (Doctoral dissertation, Indiana University).
- Limerick, P. (2017). Language contact in the US Southeast: The case of Spanish subject expression in an emerging bilingual community in Georgia. *Spanish in Context*, 14 (1), 53–77.
- Loza, S. (2021). Oral Corrective Feedback in the Spanish Heritage Language Context: A Critical Perspective. In *Heritage Language Teaching* (pp. 119-137). Routledge.
- Loza, S., & Beaudrie, S. M. (Eds.). (2021). *Heritage language teaching: Critical language awareness perspectives for research and pedagogy*. Routledge.
- Lynch, A. (1999). *The subjunctive in Miami Cuban Spanish: Bilingualism, contact, and language variability*. (Doctoral dissertation). University of Minnesota, Twin Cities.
- Martillo Viner, K. (2016). Second-generation NYC bilinguals' use of the Spanish subjunctive in obligatory contexts. *Spanish in Context*, 13 (3), 343–70.
- Martillo Viner, K. (2018). The optional Spanish subjunctive mood grammar of New York City heritage bilinguals. *Lingua*, 2010–11, 79–84.
- Martínez, G. A. (2003). Classroom based dialect awareness in heritage language instruction: A critical applied linguistic approach. *Heritage Language Journal*, 1(1), 44-57.
- Martínez-Mira, M. I. (2009). Spanish heritage speakers in the southwest: Factors contributing to the maintenance of the subjunctive in concessive clauses. *Spanish in context*, 6(1), 105-126.
- Mikulski, A. M. (2010). Receptive volitional subjunctive abilities in heritage and traditional foreign language learners of Spanish. *The Modern Language Journal*, 94(2), 217-233.
- Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and cognition*, 5(1), 39-68.

- Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and cognition*, 7(2), 125-142.
- Montrul, S. (2007). Interpreting mood distinctions in Spanish as a heritage language. *Spanish in contact policy, Social and linguistic inquiries*. Amsterdam: John Benjamins.
- Montrul, S. (2009). Knowledge of tense-aspect and mood in Spanish heritage speakers. *International Journal of bilingualism*, 13(2), 239-269.
- Montrul, S. (2014). Structural changes in Spanish in the United States: Differential object marking in Spanish heritage speakers across generations. *Lingua*, 151, 177-196.
- Montrul, S., & Bowles, M. (2009). Back to basics: Differential object marking under incomplete acquisition in Spanish heritage speakers. *Bilingualism*, 12(3), 363-383.
- Montrul, S., De La Fuente, I., Davidson, J., & Foote, R. (2013). The role of experience in the acquisition and production of diminutives and gender in Spanish: Evidence from L2 learners and heritage speakers. *Second Language Research*, 29(1), 87-118.
- Montrul, S., & Perpiñán, S. (2011). Assessing differences and similarities between instructed heritage language learners and L2 learners in their knowledge of Spanish tense-aspect and mood (TAM) morphology. *Heritage Language Journal*, 8(1), 90-133.
- Montrul, S., & Slabakova, R. (2000). Acquiring semantic properties of Preterite and Imperfect tenses in L2 Spanish. In *Proceedings from BUCLD* (Vol. 24, pp. 15-27).
- Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software*, 43, 1-14.
- Orozco, R., & Hurtado, L. M. (2020). A variationist study of subject pronoun expression in Medellín, Colombia. *Languages*, 6(1), 5.
- Otheguy, R. (2016). The linguistic competence of second generation bilinguals: A critique of “incomplete acquisition. In C. Tortora, M. den Dikken, I. L. Montoya, & T. O'Neill (Eds.), *Romance linguistics 2013: Selected papers from the 43rd linguistic symposium on Romance languages* (pp. 301–19). Amsterdam: John Benjamins.

- Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford University Press.
- Padilla, L. V., & Vana, R. (2019). Ideologies in the foreign language curriculum: Insights from textbooks and instructor interviews. *Language awareness, 28*(1), 15-30.
- Pascual y Cabo, D. (2013). *Agreement reflexes of emerging optionality in heritage speaker Spanish*. University of Florida.
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency. *Frequency and the emergence of linguistic structure, 45*(137), 10-1075.
- Polinsky, M. (2006). Incomplete acquisition: American russian. *Journal of Slavic linguistics, 191-262*.
- Polinsky, M., & Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom. *Language and linguistics compass, 1*(5), 368-395.
- Poplack, S., & Dion, N. (2009). Prescription vs. praxis: The evolution of future temporal reference in French. *Language, 557-587*.
- Poplack, S., & Levey, S. (2010). Contact-induced grammatical change: A cautionary tale. *Language and space: An international handbook of linguistic variation, 1*, 391-419.
- Quesada, M. L. (2013). The primacy of morphology in the acquisition of tense and aspect in L2 Spanish narrative structure. *Selected Proceedings of the 15th Hispanic Linguistics Symposium, 62-77*. Cascadilla Proceedings Project.
- RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- Ripley, B. (2017). MASS: Functions and datasets to support Venables and Ripley, "Modern Applied Statistics with S" (2002). R package version 7.3-47.
- Sankoff, G., & Thibault, P. (1977). L'alternance entre les auxiliaires "avoir" et "être" en français parlé à Montréal. *Langue française, (34)*, 81-108.
- Sankoff, G. (1973). Above and beyond phonology in variable rules. In C.J. Bailey & R. Shuy (Eds.). *New ways of analyzing variation in English* (pp. 44-62). Washington, DC: Georgetown University Press.

- Scrivner, O., & Díaz-Campos, M. (2016). Language Variation Suite: A theoretical and methodological contribution for linguistic data analysis. *Proceedings of the Linguistic Society of America*, 1, 29-1.
- Serafini, E. J. (2021). Assessing Students through a Critical Language Awareness Framework. *Heritage Language Teaching*, 80-97.
- Silva-Corvalán, C. (1983). Tense and aspect in oral Spanish narrative: Context and meaning. *Language*, 760-780.
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Clarendon Press.
- Silva-Corvalán, C. (2003). Linguistic consequences of reduced input in bilingual first language acquisition. *Linguistic theory and language development in Hispanic languages*, 375-397.
- Silva-Corvalán, C. (2014). *Bilingual language acquisition: Spanish and English in the first six years*. Cambridge University Press.
- Slabakova, R. (2002). Recent research on the acquisition of aspect: an embarrassment of riches?. *Second Language Research*, 18(2), 172-188.
- Schwenter, S. A., & Cacoullos, R. T. (2008). Defaults and indeterminacy in temporal grammaticalization: The ‘perfect’ road to perfective. *Language variation and change*, 20(1), 1-39.
- Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge University Press.
- Tagliamonte, S. A. (2014). Situating media influence in sociolinguistic context. *Journal of Sociolinguistics*, 18(2), 223-232.
- Thomason, S. (2020). Contact explanations in linguistics. *The handbook of language contact*, 31-49.
- Torres, L. (1989). Mood selection among New York Puerto Ricans. *International Journal of the Sociology of Language*, 79, 67-78.
- Torres-Cacoullos, R. (2012). Grammaticalization through inherent variability: The development of a progressive in Spanish. *Studies in Language. International Journal sponsored by the Foundation “Foundations of Language”*, 36(1), 73-122.

- Valdés, G. (2000). Spanish for native speakers: AATSP professional development series handbook for teachers K-16. *Spanish for native speakers*, 1-20. American Association for Teachers of Spanish and Portuguese.
- Torres Cacoullos, R. & Travis, R. (2018). *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Verkuyl, Henke J. 1972. *On the compositional nature of the aspects (Vol. 15)*. Dordrecht: D. Reidel Publishing Company.
- Weinreich, U. (1979). *Languages in contact: findings and problems*. De Gruyter Mouton.
- Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. Springer New York.
- Zentella, A. (1997). *Growing up bilingual: Puerto Rican children in New York*. Blackwell.

APPENDIX A

THE TWELVE BILINGUAL DOMINANCE SCALE QUESTIONS AND THE
SCORING PROCEDURE BY DUNN & TREE (2009)

Questions 1 and 2: At what age did you first learn Spanish _____ English _____ ?

Scoring: 0–5 yrs = +5, 6–9 yrs = +3, 10–15 yrs = +1, 16 and up = +0

Questions 3 and 4: At what age did you feel comfortable speaking this language? (If you still do not feel comfortable, please write “not yet.”)

Spanish _____ English _____

Scoring: 0–5 yrs = +5, 6–9 yrs = +3, 10–15 yrs = +1, 16 and up = +0, “not yet” = +0

Question 5: Which language do you predominately use at home?

Spanish _____ English _____ Both _____

Scoring: if one language used at home, +5 for that language; if both used at home, +3 for each language

Question 6: When doing math in your head (such as multiplying 243×5), which language do you calculate the numbers in? _____

Scoring: +3 for language used for math; +0 if both

Question 7: If you have a foreign accent, which language(s) is it in? _____

Scoring: if one language is listed, add +5 to the opposite language of the one listed; if both languages are listed, add +3 to both languages; if no language is listed, add nothing

Question 8: If you had to choose which language to use for the rest of your life, which language would it be? _____

Scoring: +2 for language chosen for retention

Questions 9 and 10: How many years of schooling (primary school through university)

did you have in: Spanish _____ English _____

Scoring: 1–6 yrs = +1, 7 and more yrs = +2

Question 11: Do you feel that you have lost any fluency in a particular language?

If yes, which one? _____ At what age? _____

Scoring: -3 in language with fluency loss; -0 if neither has lost fluency

Question 12: What country/region do you currently live in? _____

Scoring: +4 for predominant language of country/region of residence

APPENDIX B

BILINGUAL LANGUAGE PROFILE: SPANISH ENGLISH COMPETENCE BY
CARVALHO (2012-)

Competencia	(0=no muy bien 6=muy bien)
¿Cómo habla en español?	
¿Cómo habla en inglés?	
¿Cómo entiende en español?	
¿Cómo entiende en inglés?	
¿Cómo lee en español?	
¿Cómo lee en inglés?	
¿Cómo escribe en español?	
¿Cómo escribe en inglés?	

APPENDIX C

VERB FREQUENCY AND PROPORTION OF PRETERITE USE

Row Labels	Verb Frequency	Proportion
ser	117	0.27
tener	62	0.18
estar	42	0.33
empezar	37	0.95
decir	35	0.6
ir	21	0.86
hacer	19	0.79
vivir	18	0.67
venir	15	0.93
nacer	14	1
trabajar	13	0.62
dar	12	0.5
casarse	12	0.83
haber	12	0.17
llegar	10	0.9
durar	9	1
quedarse	8	1
ver	8	0
poder	8	0
estudiar	8	1
salir	8	0.75

aprender	8	0.88
querer	7	0.29
gustar	6	0.83
entrar	6	0.67
morir	6	1
conocer	5	0.8
ayudar	5	0.2
terminar	5	0.8
moverse	5	0.8
mirar	4	0.5
graduarse	4	1
pedir	4	0.5
seguir	4	0.75
tocar	4	1
dejar	4	0.5
llevar	4	0.5
parecer	3	0
considerar	3	1
creer	3	0
levantarse	3	0.67
cambiar	3	0.67
volver	3	1

venirse	3	1
hacerse	3	1
crecer	3	1
tomar	3	1
meter	3	0.67
quedar	3	0.67
andar	3	0
pensar	3	0.67
regresar	3	1
batallar	2	1
sentarse	2	0.5
subirse	2	0.5
ocurrir	2	0.5
esforzar	2	1
agarrar	2	0.5
asistir	2	0.5
ponerse	2	0.5
saber	2	0.5
sufrir	2	1
carecer	2	0
decidirse	2	1
caer	2	0.5

sacar	2	0.5
mandar	2	0.5
gritar	2	1
adaptarse	2	0
sentirse	2	0
enseñar	2	0.5
hablar	2	1
ocupar	2	0
lastimar	2	1
importar	1	0
preguntar	1	1
convivir	1	1
huir	1	1
arreglar	1	1
platicar	1	0
criar	1	1
pertenecer	1	0
compartir	1	1
retirarse	1	1
bajar	1	1
requerir	1	0
adorar	1	0

pasar	1	1
darse	1	1
participar	1	1
llorar	1	0
rentar	1	0
quebrar	1	1
festejar	1	1
llamarse	1	0
fallecer	1	1
bautizar	1	1
existir	1	0
irse	1	1
exigir	1	0
invadir	1	1
parar	1	1
aceptar	1	0
olvidar	1	1
decidir	1	1
reir	1	0
suceder	1	1
escribir	1	0
voltear	1	0

enviar	1	0
costar	1	1
sentir	1	0
convertir	1	1
necesitar	1	0
provener	1	0
enfocarse	1	0
llegamos	1	1
encontrar	1	1
llamar	1	1
regañar	1	1
buscar	1	1
emigrar	1	1
levantar	1	0
recibir	1	1
usar	1	0
dividir	1	1
arrepentirse	1	1
despertarse	1	1
temer	1	0
desanimarse	1	1
introducir	1	1

recargarse	1	0
imprimir	1	1
dedicarse	1	0
acabar	1	0
tratar	1	0
meterse	1	1