UNIVERSITY OF CALIFORNIA, SAN DIEGO

Traversing the Wall: A Study of Language Contact among Heritage and Immigrant Speakers of Spanish in the Tijuana-San Diego Border Area

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Linguistics

by

Rodolfo Mata

Committee in charge:

Professor John Moore, Chair
Professor Gabriela Caballero
Professor Grant Goodall
Professor Kathryn A. Woolard
Professor Ana Celia Zentella

2016

The dissertation of Rodolfo Mata is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2016

DEDICATION

For Michell and Gloria

For all the families separated by the border

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Thank you to my advisors Grant Goodall and John Moore. To the faculty, the administrative and IT staff, the academic coordinators, and the language lab staff at the UCSD Department of Linguistics. To my family, my friends, my colleagues in the LLP and my students. To my dissertation committee. To the families in Tijuana and San Diego whose life stories are the rich data for this project. To everyone who participated in the Border Spanish Project.

VITA

2002	B.A. in English and Linguistics University of Texas at El Paso
2005	M.F.A. in English, Creative Writing University of Alaska, Fairbanks
2011	M.A. in Linguistics University of California, San Diego
2016	Ph.D. in Linguistics University of California, San Diego

ABSTRACT OF THE DISSERTATION

Traversing The Wall: A Study of Language Contact
among Heritage and Immigrant Speakers of Spanish
in the Tijuana-San Diego Border Area

by

Rodolfo Mata

Doctor of Philosophy in Linguistics

University of California, San Diego, 2016

Professor John Moore, Chair

In the study of emerging varieties of Spanish in the United States, the Tijuana-San Diego border area presents a unique opportunity for the study of language contact in that English and two varieties of Spanish (U.S. and Mexico) are in constant contact

with one another. In the San Diego area we find two types of Spanish native speakers, corresponding to two generations: a heritage group that is English-dominant and an immigrant group that is Spanish-dominant. Tijuana speakers represent monolingual controls that are the closest point of reference, linguistically and demographically, to immigrant and heritage speakers in San Diego. In a fieldwork study of 22 families (11 on each side of the border) that consists of naturalistic spoken data in a conversational setting, I focus on two linguistic features of Border Spanish: the use of the subjunctive and the use of fillers. With respect to the subjunctive, heritage speakers exhibit an attenuation of the imperfect subjunctive in optional contexts and an increase of the imperfect subjunctive outside of subjunctive contexts. I propose that this difference is due to a pattern of interactions and exposure to Spanish, unique to heritage speakers, that coincides with the onset of formal schooling in English and a gradual shift to English dominance that prevents heritage speakers from fine-tuning their use of the subjunctive in certain contexts, leading to the observed effects.

With respect to fillers, heritage speakers complement their system of Spanish fillers with English fillers. In spite of not being English dominant, some immigrant speakers begin to use English fillers with limited functions when compared to heritage speakers. These indirect transfer effects in the use of fillers may be due to San Diego speakers' highly-variable exposure and interactions in both Spanish and English. Whereas heritage speakers may transfer the filler 'so' directly from English, immigrant speakers' transfer may be the result of interaction both with heritage speakers and exposure to English. This dissertation explores the role that exposure and interactions

in Spanish and English, English-language dominance, and the lack of formal education in Spanish play in contact-induced language change in the Spanish spoken in San Diego.

Chapter 1 Introduction

1.1 Overview

The focus of this dissertation is the Spanish spoken in the Tijuana-San Diego border area, specifically the effects of language contact with English on the speech of immigrant and heritage populations in San Diego. When speaking of language contact I refer to the theoretical construct set forth by Otheguy and Zentella (2012) in their study of Spanish in New York wherein language contact is used "to explain the differences between bilingual and reference lects associated with the same language". In the Tijuana-San Diego border area, the bilingual lects are those spoken by heritage speakers in San Diego and the reference lects by the speakers in Tijuana and immigrant speakers in San Diego. Of particular interest to this dissertation is the question of how a heritage lect is created in the San Diego area. In other words, what are the characteristics of heritage Spanish in San Diego? How is heritage Spanish in San Diego different the reference lects spoken in the border area? And what are the possible causes of the emergence of a heritage lect in San Diego?

To answer these questions I focus on the subjunctive and the use of fillers[1]. The subjunctive has been identified as a particularly vulnerable area of Spanish in contact with English. Silva-Corvalán (1994) proposes that although the subjunctive undergoes simplification in monolingual varieties of Spanish – following a general tendency among Romance languages – a language-contact situation may accelerate its

[1] Fillers in Spanish include discourse markers such as *pues*, *bueno*, *entonces*, *eh*, and *este* among others. I present a full discussion in chapter 6 of this dissertation.

simplification or the loss of some of its features. In data collected from monolingual children in Mexico City, Blake (1983) found that the subjunctive is acquired gradually from the ages of 3 until approximately ten[2]. For second-generation San Diego speakers, this is the age period that coincides with the onset of schooling in English as well as with a shift in exposure and interactions in English. In spite of being English-dominant as adults, heritage speakers in San Diego are highly proficient and highly functioning in Spanish. They differ from their parents and from the Tijuana speakers in their lack of access to formal schooling in Spanish and in a radical shift during childhood to English as the dominant language. While differences in subjunctive use among heritage Spanish-speaking populations have been analyzed as a result of incomplete acquisition or language attrition (Silva-Corvalán 2003; Montrul 2009), I examine the differences in use of subjunctive, and their causes, by looking at formal schooling in English and a unique set of interactions and exposure to Spanish and English that is characterized by English dominance.

In regards to fillers, the borrowing of English fillers into Spanish in a language-contact situation has been documented in studies by Torres & Potowski (2008), Lipski (2005), and Silva-Corvalán (1994), among others. Since fillers are particularly prone to borrowing, Fuller (2001) has pointed out that one of the results of intense contact between two languages is a mixed system in which the fillers from both languages are used. The donor language is the one that is pragmatically dominant and the borrowing of fillers tends to be gradual (Matras 1998). For the heritage speakers, English is the

[2] Blake (1983) notes that some subjunctive structures do not reach mastery until adulthood.

pragmatically-dominant language since they borrow the English fillers into Spanish. Immigrant speakers in San Diego, however, also exhibit English fillers (albeit one) as part of their speech production even though they are Spanish dominant – this positions the contact variety of Spanish as the pragmatically dominant lect in their community. The borrowing of English fillers into the Spanish of immigrant and heritage speakers in San Diego raises some valid questions about the effects of interaction and exposure to two pragmatically-dominant languages: English, and a variety of U.S. Spanish that has integrated English fillers.

As I mentioned, both the use of subjunctive and fillers have been identified as vulnerable areas in Spanish in contact with English in the United States. Differences in use of subjunctive and fillers among heritage populations has been reported in various studies[3], but the San Diego area remains understudied regarding these two linguistic features. In looking at the possible causes for the differences in use of subjunctive and fillers, I set aside questions of incomplete acquisition and language attrition. Incomplete acquisition, as a theoretical construct, refers to an incomplete grammar in the native language that heritage speakers never fully acquire and which weakens as socialization in the non-heritage language intensifies (Benmamoun, Montrul, & Polinsky 2010). The implication is that any features of the heritage language that have been delayed in language development never reach native or target-like levels. Benmamoun, Montrul, & Polinsky (2010) point to features that prove to be problematic to an adult heritage

[3] For a more thorough discussion of these studies on the use of subjunctive and fillers in U.S. varieties of Spanish, see chapters 5 and 6 of this dissertation.

speaker and to a five-year-old speaker as likely indicators of incomplete acquisition of this feature.

Otheguy (2016) rejects the notion of incomplete acquisition to describe the acquisition process of heritage speakers since it is impossible to determine what exactly would constitute complete acquisition. Moreover, Otheguy criticizes the concept of a target that heritage speakers should try to attain since usually this so-called target is based on a standardized variety of a grammar that is created in an environment much different from the one where heritage speakers grow up – namely, one where formal education grants a speaker familiarity with standardized features of a language. Work by Otheguy & Zentella (2012) and Otheguy (2016) has described the grammar of heritage speakers of Spanish in the United States as one that has evolved differently but which is not incomplete. Differences in the Spanish of heritage speakers do not constitute errors or incompleteness but rather they represent dialectal differences and intergenerational changes that may be "accelerated by language contact" (Otheguy 2016).

Even though Montrul (2010) notes that incomplete acquisition can be more clearly evidenced in simultaneous bilinguals than in sequential bilinguals[4], I propose that the heritage population under study in this dissertation does not possess an incomplete grammar but rather a different one that is typical of what Otheguy (2016) identifies as normal language transmission in a bilingual or multilingual setting, wherein

[4] The San Diego second-generation speakers can be considered sequential bilinguals in that they all report Spanish as their first language and English as their second.

heritage speakers without access to literacy in Spanish grow up with "models of popular speech" (Otheguy 2016) in Spanish-speaking U.S. communities and not with a standardized Latin American model.

Language attrition is even more problematic than the concept of incomplete acquisition in speaking about the grammar of the heritage speakers in San Diego that are part of this study. Benmamoun, Montrul, & Polinsky (2010) note the difficulty of ascertaining whether a feature has been acquired in childhood[5] and then lost later in life, and whether attrition even occurs in heritage language acquisition given the great degree of individual speaker variation. Only longitudinal data collected before the onset of education in English for the San Diego heritage speakers – and then compared with the current data collected in early adulthood – could determine if use of subjunctive and fillers has changed over time and if this difference in use reveals a loss of features that could be considered language attrition. Given that no such data was collected for this study, this question is outside the scope of this dissertation

In the following section I provide an overview of the language-contact situation in the border area.

1.2 Language Contact: Exposure and Interactions in Spanish and English

I focus on exposure to Spanish and English and interactions in these two languages as a way of looking at language contact phenomena as the source of language change, which in the border area is manifested in the speech of San Diego speakers. The

[5] By the age of 4 or 5 (Benmamoun, Montrul, & Polinsky 2010).

heritage speakers of Spanish in this study experience a great degree of variability in exposure and use of Spanish from early childhood. The input they are exposed to is not always in Spanish but rather varies greatly from speaker to speaker (depending on external factors such as order of birth, onset of acquisition in English, language interactions within the family, etc.). The onset of formal schooling in English for heritage speakers signals a crucial shift in terms of language dominance for two reasons:

1. Social interactions and exposure to both languages begin to shift for heritage speakers in that they gradually become more English-dominant outside the home[6]. Spanish, which all heritage speakers in the study report as their first language, is used primarily in interactions with family and friends, and later on in life to some extent at their place of employment. None of the heritage speakers in the study are married, therefore there is no data to determine the language of interactions with a spouse, extended family, and children. With the beginning of schooling in English, the language experience of heritage speakers changes towards English dominance.
2. At the onset of education, monolingual speakers of Spanish in Tijuana begin to receive input within a language framework that is ruled by normative grammatical tenets. Heritage speakers in San Diego do not always have access to such a language framework, and if they do it is at best limited to a couple of years of bilingual education or Spanish classes (L2 or Heritage)

[6] Heritage speakers may become English dominant inside the home if they are third generation or if they have siblings who only communicate with them in English.

> for some of them. Not only do heritage speakers become English dominant, but also the input they receive in Spanish is different from a monolingual Tijuana speaker or an immigrant San Diego speaker due to a lack of formal schooling in Spanish.

As such, the two most relevant aspect of the language-contact situation for heritage speakers are English dominance and lack of formal schooling in Spanish.

None of the immigrant speakers in San Diego become English dominant but their exposures and interactions exhibit some degree of variability. For instance, immigrant speakers for at least 18 years have had reduced contact with non-contact variety of Spanish in Mexico. They have since primarily been exposed to a San Diego lect that may have undergone some changes as a result of language contact, and for immigrant speakers this San Diego lect represents the dominant language in their community. They also have exposure to English and to other Spanish lects (contact and non-contact) not only in person but also through the media. The most salient factors of the language contact situation for immigrant speakers in San Diego in this study are the variability in the languages and lects of their interactions and exposure, and their place in a community where the U.S. lect of Spanish is the dominant language. In chapter 3 I provide detailed description of the interactions and exposure to Spanish and English for all populations.

1.3 Four Populations in the Border Area

This study focuses on 22 families in the San Diego-Tijuana border area, eleven on each side of the border. For the purpose of this study, a family is defined as one parent (the first generation) born in Mexico and one of his/her adult children (the second generation). Based on location, there are two major groups of speakers in the border area: San Diego and Tijuana. To include location and generation when referring to each of the populations throughout the dissertation, the four groups of speakers will be labeled as follows in this study:

1. TJ1 for the Tijuana first generation
2. TJ2 for Tijuana second generation
3. SD1 for San Diego first generation
4. SD2 for San Diego second generation.

The TJ1 and TJ2 populations are sometimes referred to as the Tijuana speakers with no generational distinction. The SD1 group is sometimes referred to as the immigrant group/speakers. I also refer to the SD2 group as the heritage group/speakers. Additionally, any reference to non-heritage speakers refers to the TJ1, TJ2, and SD1 groups combined.

The following three subsections present a general description of each group and provide an overview of their language contact situation. For each group I also outline the major points of interaction and exposure to Spanish and English. In chapter 3 of this dissertation I provide an in-depth look at the exposure and interactions in Spanish and English for all populations in the border.

1.3.1 Heritage Speakers in the United States

A heritage speaker is someone who early on was exposed to and acquired a language that is different from the dominant language of the area where s/he lives. In the United States a heritage speaker is someone who is raised in a home environment where a language other than English is spoken, and who is bilingual in English and in the home language. Polinsky (1997) refers to the heritage speaker's home language as the baseline language, or control language, one which the speaker acquires directly from his/her parent(s) and in which the speaker rarely has access to formal schooling. The baseline language for a heritage speaker is not the same as the standard variety of the same language that was acquired by competent, educated speakers in their home countries. The baseline is what Polinsky (2008) refers to as the language that heritage speakers use with their families and with their communities. In noting the difficulty in defining exactly what constitutes a heritage language speaker, Carreira (2004) establishes the following criteria as possible determining factors: the place the speaker has in the community where the baseline is spoken, the speaker's connection to his/her heritage language and culture, and the speaker's proficiency in the heritage language. It is this third factor, a heritage language speaker's proficiency, that presents a challenge when it comes to defining and categorizing heritage language speakers due to the instability of the baseline language.

Valdés' (2000) definition of a heritage language speaker centers only on proficiency in the baseline language. For Valdés, a heritage language speaker grew up in a family environment where a language other than English was the primary language;

he/she has some degree of oral and listening proficiency in the heritage language, and is to a certain extent bilingual in the baseline language and English. Some of the sociolinguistic factors that impact proficiency in the heritage language, as they apply particularly to heritage speakers of Spanish in the United States, include, as Carreira (2004) mentions, a very strong and popular Spanish media that highlights not only the prevalence of Spanish in the United States but also the great extent to which Spanish is an essential part in the lives of Latinos in the United States. The baseline proficiency of a heritage speaker of Spanish in the U.S. is partly determined by the exposure the speaker has to Spanish in media (present most ubiquitously in television and radio among Latino communities) and in social contexts such as school, the workplace, and church services and communities, where Spanish is widely used.

The diversity of Latino populations in the United States also contributes to the instability of the core language since heritage speakers are rarely exposed to the standard variety of Spanish, and in some cases their exposure comes from interacting with family members and friends who themselves may be heritage speakers of Spanish. In her study of heritage speakers of Portuguese in the United States, Silva (2008) draws a parallel between her subjects and Silva-Corvalán's (1994) Spanish-speaking subjects when she notes that exposure to a non-standard variety of the baseline language as well as contact with English results in a core language that is not altogether stable due to its innovative forms. Competent speakers of the language, notes Polinksy (2008), have also been exposed to different varieties of the language but because they have received formal

education in the standard variety, their language input results in a less-varied baseline than heritage speakers.

In addition to the difficulty in defining what exactly a heritage speaker is, categorizing heritage speakers presents another difficulty given that the irregularity of the baseline language prevents them from being considered a homogenous group. The one trait that most heritage speakers of Spanish share is that, in their language histories, Spanish (the L1) starts out as the primary language until the age of schooling begins, when English (the L2) replaces Spanish as the predominant language. After this point, the multitude of variants makes it impossible to categorize heritage speakers of Spanish into discrete categories. Some of these speakers' exposure to Spanish as children might have been insufficient at best while others may have acquired it from fully competent parents with a very stable baseline. For others, the shift from Spanish to English may have been very rapid and Spanish was never acquired completely. Montrul (2002) notes that considering that many of these bilinguals feel more comfortable using English than Spanish – since they seemed to never acquire full competence in Spanish – in their Spanish proficiency they more closely resemble second-language learners than bilinguals. Polisnky (2008) acknowledges the difficulty in sub-categorizing heritage speakers. She suggests that looking at factors such as lexical knowledge, mean length of utterances, and speech rate in combination is a more accurate form of determining heritage speaker proficiency than looking solely at sociolinguistic factors.

1.3.2 Heritage Speakers (SD2)

In this study, the Heritage group (SD2) consists of the second-generation speakers in San Diego who are the children of the Immigrant (SD1) group. All eleven speakers in this group were born in the United States and raised in the San Diego area. All report Spanish as their first language and acquisition of English at the onset of education. All speakers in this group are now English dominant. Their ages range from 18 to 25 at the time of the interview. What is worth underscoring for this group is the high variability of exposure to and use of Spanish in their daily lives beginning at the onset of schooling.

1.3.3 Immigrant Speakers (SD1)

The Immigrant group (SD1) consists of eleven first-generation San Diego speakers who are immigrants to the San Diego area from various parts of Mexico and who have lived in the United States for at least 18 years. This group's first language is Spanish and some have acquired English to varying degrees in adulthood as a result of living in the United States. Their ages range from 41 to 60 at the time of the. Their socioeconomic status is primarily working class and their educational level similar to the TJ1 group. In terms of exposure to English all subjects have had exposure to English as a result of living in the United States. However their levels of proficiency in English vary greatly. This variation in use of English is the distinguishing feature of this population.

1.3.4 Tijuana Speakers (TJ1 and TJ2)

The Tijuana group is the native control group and it is subdivided into two groups: eleven parents (TJ1) and eleven children (TJ2). The first Tijuana group (TJ1) consists of speakers in Tijuana, Mexico, who have never lived in the United States and whose first and dominant language is Spanish. Their ages range from 36 to 69 years of age at the time of the interview. Within the group, there is variability in socioeconomic status and educational level to an extent that is comparable to the SD1 group. Unlike the San Diego populations, many in the TJ1 group report limited exposure to English outside of the media. In terms of migration to the border area, only two of the speakers report being from the Tijuana area whereas the rest come from elsewhere in Mexico and have spent at least 15 years in Tijuana.

The second Tijuana group (TJ2) consists of the second-generation speakers in Tijuana who are the children of the first group (TJ1) and, who like their parents, have never lived in the United States and whose first and dominant language is Spanish. Their ages range from 18 to 30 at the time of the interview. While their socioeconomic status is the same as their parents, since most of the second-generation Tijuana subjects live at home, there is less variation in their educational level. What distinguishes the second-generation Tijuana speakers from their parents is a higher educational level across the board. There is great variation in the speakers' exposure to English. All speakers were born and raised in Tijuana, or born elsewhere in Mexico but raised in Tijuana from an early age.

1.4 Relevance of the Present Study

This dissertation contributes to a growing body of research that seeks to investigate the linguistic properties of the Spanish spoken in the U.S. The data gathered consists naturalistic conversational speech that reflects how speakers use Spanish in an everyday setting. The variety of Spanish spoken in the Tijuana-San Diego border area remains largely understudied[7] when compared to other varieties of Spanish in the United States. Furthermore, the border area presents a unique sociolinguistic environment in that speakers on both sides of the border have immediate access to both varieties of Spanish. In the interviews that are part of the corpus for the present study, speakers on both sides of the border report that up to recently there had been a constant back and forth movement between people in Mexico and the U.S. Speakers in Tijuana report having family members in the San Diego and Los Angeles areas that they remain in personal contact with. A number of immigrant speakers in San Diego maintains close ties to Tijuana since for some of them Tijuana represented the intermediate point of migration between the interior of Mexico and the United States. Some San Diego speakers report having family members in Tijuana that they visit frequently or who visit them in the U.S. This back-and-forth between the populations due to their close geographic proximity – in addition to both having access to their respective countries' Spanish-language media – creates a sociolinguistic region that is ideal for the study of language contact.

[7] See work by Zentella (2009) and Relano Pastor (2007)

A number of studies on Spanish in the U.S. focus on comparisons between heritage and immigrant speakers, or heritage and Latin-American raised speakers. The present study is unique in that it takes into account not only heritage and immigrant speakers but also speakers in Tijuana whose sociodemographic profile is similar to the immigrant populations in San Diego. The Tijuana speakers serve not only as native controls but they also allow for a more balanced three-way comparison among populations given the aforementioned sociodemographic similarities between the first generations on both sides of the border.

1.5 Dissertation Overview

The overarching research questions that guide this research project are the following:

1. What are the characteristics of heritage speech in the Spanish of San Diego second-generation speakers as it pertains to use of subjunctive and fillers?
2. What are the possible causes of the emergence of this heritage lect in San Diego?

To answer these two questions I organize the rest of this dissertation as follows:

Chapter 2 presents the methodology for recruiting speakers, gathering the data, carrying out the interviews, and coding and analyzing the data. In this chapter I also provide a detailed description of the sociodemographic data for each of the four groups of speakers under study.

Chapter 3 provides a detailed description of the language experience of the four populations under study. I focus on the speakers' interactions and exposure to both Spanish and English as the crucial part of their language experience. I determine that the San Diego groups are characterized by their variable exposure and interactions in both languages.

Chapter 4 presents the results of two different methods of assessing speakers' proficiency to determine the extent to which the populations are different in regards to lexical proficiency and speech rate. Additionally, I present the results of an audio perception study to determine if monolingual Mexican speakers outside of the border area can recognize differences among the groups under study.

Chapter 5 presents the study on the subjunctive system. I provide a study of the subjunctive mood in Spanish and an overview of previous work that has been dedicated to the study of the subjunctive mood in populations in the U.S. I then provide the results of a logistic regression analysis of the subjunctive mood using a combination of language internal and external (sociodemographic) factors as predictors of subjunctive use.

Chapter 6 presents the study of fillers. I first provide an overview of the study of fillers among speakers of Spanish in the U.S. I then provide the results of quantitative and qualitative analyses that reveal three distinct ways that speakers use fillers in the border area, particularly as it applies to the presence and use of English fillers among the San Diego populations.

Chapter 7 provides a summary of the major findings on subjunctive use and the systems of fillers in the border.

Chapter 2 Methodology & Populations

This study is part of a larger study at the University of California, San Diego: the Border Spanish Project[1]. This is a project that seeks to analyze the theoretical and sociolinguistic issues that are present in the varieties of Spanish spoken in the San Diego-Tijuana area. In this chapter I describe the methodology for performing and completing the study (section 2.1) as well as the populations under analysis (section 2.2).

2.1 Methodology

2.1.1 Data Gathering: The Interviews

The data were gathered in the form of personal interviews in a conversational setting to create a corpus of naturalistic spoken data. Families in the Tijuana area were recruited through various means: some are personal friends of one of the researchers; others are personal friends, family, and acquaintances of two undergraduate researchers who are originally from Tijuana; and others are members of a Christian church whose pastor offered members of the congregation the opportunity to be part of the study. The families in the U.S. are primarily from the National City area and most were recruited through a science teacher at Sweetwater High School in National City, a largely Hispanic community in southern San Diego County. According to the U.S. census, in

[1] UCSD IRB Project Number 131385

2010 (around the time the interviews were recorded), the population in National City was 63% Hispanic.

A total of 41 interviews were recorded in Tijuana by either the researchers or the two undergraduate assistants from Tijuana. Ultimately, 22 of the interviews were selected for analysis given that the 19 discarded were determined to be too short in length (less than 10 minutes). In San Diego, 24 interviews were recorded and 22 were selected for analysis. As it was the case with the discarded interviews in Tijuana, the 2 interviews discarded in San Diego were due to insufficient length. Each set of 22 interviews corresponds to 11 families, where a family is defined as one parent and one child over the age of 18. In sum, 11 families were interviewed on each side of the border for a total of 22 families, or 44 speakers.

To identify the subjects in this study, I employ a labeling system based on their geographic and generational variants. Thus, the speakers labeled TJ-A1 and TJ-A2 refer to Tijuana speakers, both members of family A, first and second-generation respectively. Any examples from the interview data are labeled according to this system. In Appendix 2, I provide a list of all the speakers involved in the study, their gender, age, and place of birth, and their current place of residence. Gaps in the ABC labeling sequence are the result of either incomplete interview sets, only one of the family members being available for the interview, or an interview set being discarded due to insufficient length.

The Tijuana interviews took place either at the homes of the subjects, in a semi-private gathering room at a Christian church, or at the home of one of the researchers.

These interviews were conducted primarily by the researchers with the exception of four interviews (two in Tijuana and two in San Diego) that were conducted by two U.C. San Diego undergraduate students. In San Diego, six of the National City families were interviewed at Sweetwater High School either by the undergraduate assistants or by the researchers, while the remaining five families were interviewed by the researchers at the homes of the subjects. All interviews were digitally recorded and stored as WAV files. The audio files where then optimized with Audacity and transcribed by the researchers, with the exception of two families in Tijuana and two in San Diego whose transcriptions were done by two undergraduate research assistants and a linguistics graduate student. The transcription included of all spoken data, filled pauses, interruptions by other speakers, and codeswitches into English. All transcriptions were personally checked for accuracy.

Each interview was divided in three parts. The first part centered on sociodemographic information, the second part on language use, and the third on the speakers' life experiences. The sociodemographic information consisted of name, age, place of origin, number of years in the border area, occupation, and educational level. For speakers in Tijuana, language use refers only to any use of English either at school or work. For San Diego speakers, language use pertains to the extent to which they use Spanish and English at work, at home with their families, with friends and acquaintances, and with strangers. Additionally, San Diego speakers were asked to report the extent of exposure to and use of Spanish and English in writing, reading, television and movies, and listening to music. The last part of the interview, which

centered on life experience, asked subjects to describe the place where they are from and the place where they currently live, narrate about one of the happiest times in their lives, hypothesize on how they imagine their lives to be different were they living on the side of the border, express wishes and advice for future generations, and give their opinion about the Spanish spoken in the border area. In Appendix 1 I provide a sample text of the questions for the Tijuana and San Diego area interviews.

The interviews were conducted in Spanish and the subjects were informed that the purpose of the study was to study the Spanish in the border area as well as people's histories and connection to the language. Even though codeswitching was not discouraged, only three of the SD2 speakers codeswitched during the interview. However, for these three speakers the switches were brief and minimal, and the interactive and communicative situation was kept in Spanish.

2.1.2 Coding

The interviews were transcribed and coded for verb form and for sociodemographic information for each speaker. The text for each interview transcription was broken down so as to have one sentence per line. Each sentence was then entered into an Excel spreadsheet and its verb coded. Appendix 3 provides a summary and a description of the codes for both the verb form and sociodemographic codes.

2.1.3 Data Analysis

The interviews were analyzed for proficiency (see chapter 4), use of fillers (chapter 5), and use of the subjunctive mood (chapter 6). The methodologies for each of the three analyses are described within their particular chapter.

2.2 Populations

In the sections that follow (2.2.1-2.1.8) I provide a sociodemographic overview of the four groups of participants.

2.2.1 Gender

For all groups, gender was balanced. Of the 44 subjects, 22 are female and 22 are male. The gender breakdown per group is presented in Table 2.1.

Table 2.1 Gender Breakdown per Group

Group	Females	Males
TJ1	5	6
TJ2	6	5
SD1	6	5
SD2	5	6

2.2.2 Age

There were no age requirements for first-generation participants. Second-generation participants had to be 18 years of age at the time of the interview. In Table 2.2 I present the ages of all the participants sorted per group in ascending order from

youngest to oldest. The mean age is provided for each group as well as the standard deviation within the group.

Table 2.2 Age of Participants. Sorted in ascending order by age.

TJ1		**TJ2**		**SD1**		**SD2**	
	age		age		age		age
TJ-T1	36	TJ-O2	18	SD-C1	41	SD-A2	18
TJ-Q1	37	TJ-Q2	18	SD-B1	46	SD-B2	18
TJ-P1	39	TJ-R2	18	SD-K1	47	SD-D2	18
TJ-U1	40	TJ-T2	18	SD-I1	49	SD-F2	18
TJ-O1	43	TJ-U2	18	SD-E1	50	SD-H2	18
TJ-S1	44	TJ-V2	18	SD-F1	50	SD-I2	18
TJ-R1	46	TJ-P2	19	SD-J1	50	SD-K2	20
TJ-V1	54	TJ-S2	21	SD-L1	50	SD-L2	20
TJ-M1	57	TJ-B2	22	SD-D1	52	SD-M2	20
TJ-F1	62	TJ-M2	23	SD-H1	52	SD-C2	21
TJ-L1	69	TJ-L2	31	SD-M1	60	SD-J2	25
Mean Age	47.9		20.36		49.7		19.5
Standard Deviation	10.9		3.9		4.6		2.1

The large age ranges for the first-generation groups are to be expected given that there were no age requirements. Every effort was made to minimize the age range for the second-generation groups so that they would represent a more homogeneous group. This proved more challenging in Tijuana, where the oldest speaker in the second-generation is aged 31.

2.2.3 Years in the Border

Second-generation speakers in San Diego were required to have been born in the United States, raised in the San Diego area, and not lived in Mexico. Second-

generation speakers in Tijuana were required to have been raised in the Tijuana area. For first-generation speakers, those in San Diego were required to have immigrated to the United States as adults from somewhere in Mexico. For Tijuana first-generation speakers there were no requirements other than currently be living in Tijuana. In Table 2.3 I present the number of years living in the border area (YB) for each speaker next to their age. For San Diego speakers this number indicates the number of year living in the United States.

Table 2.3 Years in the Border. YB indicates the number of years the speaker has lived in the border area.

TJ1				TJ2				SD1				SD2		
	age	YB			age	YB			age	YB			age	YB
TJ-T1	36	15		TJ-T2	18	15		SD-C1	41	18		SD-A2	18	18
TJ-Q1	37	15		TJ-R2	18	16		SD-B1	46	18		SD-B2	18	18
TJ-U1	40	15		TJ-U2	18	16		SD-I1	49	20		SD-D2	18	18
TJ-R1	46	16		TJ-O2	18	18		SD-E1	50	20		SD-F2	18	18
TJ-S1	44	25		TJ-Q2	18	18		SD-H1	52	21		SD-H2	18	18
TJ-M1	57	35		TJ-V2	18	18		SD-F1	50	22		SD-I2	18	18
TJ-P1	39	39		TJ-P2	19	19		SD-L1	50	24		SD-K2	20	20
TJ-O1	43	43		TJ-M2	23	20		SD-K1	47	25		SD-L2	20	20
TJ-L1	69	48		TJ-S2	21	21		SD-M1	60	25		SD-M2	20	20
TJ-V1	54	51		TJ-B2	22	22		SD-J1	50	28		SD-C2	21	21
TJ-F1	62	56		TJ-L2	31	31		SD-D1	52	30		SD-J2	25	25

2.2.4 Place of origin

In Table 2.4 I present the places of origin for all speakers. For the Tijuana speakers, I identify the state and not the city of origin except for Tijuana. For San Diego, for SD1 speakers I identify only the state. This includes Baja California Norte (B.C.N.) with one speaker from Tijuana and one from Mexicali. For SD2 speakers, I also identify

the city and not the state to differentiate the two speakers born in Los Angeles and not in San Diego.

Table 2.4 Place of Origin for All Speakers

TJ1		**TJ2**		**SD1**		**SD2**	
Tijuana	2	Tijuana	7	Jalisco	4	San Diego	9
Sonora	2	Coahuila	2	Sinaloa	2	Los Angeles	2
Coahuila	2	Sonora	1	B.C.N.	2		
Guanajuato	1	Morelos	1	Chihuahua	1		
Jalisco	1			Queretaro	1		
DF	1			Oaxaca	1		
Morelos	1						
Colima	1						
Total	11		11		11		11

2.2.5 Current residence

All 22 Tijuana speakers reside in Tijuana. Of the San Diego speakers, 20 reside in National City and two (SD-L1 and L2) in Lemon Grove.

2.2.6 Socioeconomic status

The majority of speakers on both sides of the border come from a working class background. Because all of the second-generation speakers live with their parents, their socioeconomic status is the same for both generations. This data is presented in Table 2.5.

Table 2.5 Socioeconomic Status

	TJ1	**TJ2**	**SD1**	**SD2**
Working class	8	8	9	9
Professional	2	2	1	1
Business owner	1	1	1	1
Total	11	11	11	11

2.2.7 Occupation

Eight out of 11 of the first-generation speakers on both sides of the border are working class or homemakers. Similarly, the majority of second-generation speakers are students. The occupation for each of the speakers is presented in Table 2.6.

Table 2.6 Occupation

Profession	**TJ1**	**TJ2**	**SD1**	**SD2**
Working class	5	1	8	2
Professional	1	1		
Business owner	1		1	
Homemaker	3		1	
Retired	1			
Office Worker		1	1	
Student		8		9
Total	11	11	11	11

2.2.8 Education

The education level of the second-generation speakers is generally higher than that of their parents. All second-generation speakers in Tijuana are either in high school or in college, with one speaker having completed college. All SD2 speakers have

completed high school at the time of the interview, with six of them attending college and one a graduate program. The complete breakdown of the speakers' education levels is presented in Table 2.7.

Table 2.7 Education

Education	**TJ1**	**TJ2**	**SD1**	**SD2**
Some primary				
Primary finished	2		1	
Some middle	1		1	
Middle finished	4		2	
Some HS/in HS		6	1	
HS finished	2		3	4
Some college/in college		4	1	6
College finished	1	1	2	
Some grad/in grad				1
Grad finished	1			
Total	11	11	11	11

Along with the information presented in the previous section on occupation (Table 2.7), the education level data demonstrates that there is a pattern of socioeconomic upward mobility on the part of the second generations on both of sides of the border. This becomes relevant in the type of language experience that the speakers have in relation to their education and employment. It may indicate that while second-generation speakers parallel each other's upward trajectory, their language experiences may go in different directions: as TJ2 speakers pursue higher education in Spanish, their command of different registers and vocabularies may surpass their parents'; SD2 speakers pursue

higher education in English and become more dominant in this language as their interactions and exposure to Spanish may decrease along with their Spanish proficiency.

2.3 Comparing Heritage Speakers to Monolingual Speakers

Otheguy (2016) points out that studies that compare Latin American-born controls to U.S. heritage speakers of Spanish put the heritage speakers at great disadvantage and ultimately do not provide a valid comparison between the two linguistics systems since heritage speakers do not always grow up in a language environment that can lead to a comparable grammar system to the Latin American-born controls. According to Otheguy (2016), second-generation U.S. speakers of Spanish have a language experience that is vastly different from Latin American-born controls who may be highly educated in Spanish in their home countries and whose day-to-day interactions are completely in Spanish. In the present study I address these differences between Mexican-born and U.S.-born speakers in the following ways:

1. The SD2 speakers are highly functional and demonstrate high proficiency in Spanish (see chapter 4). Although they are English dominant outside the home, their interactions with both of their parents are completely or mostly in Spanish. Moreover, a number of their interactions (with friends, with other family members, etc.) take place in Spanish as well as in English (see chapter 3). The SD2 speakers in this study do not represent a low-proficiency group. Rather, their language experience in a Spanish-dominant home in the border

area has allowed them to form a grammar that allows them to perform at a level that is comparable to non-heritage speakers.

2. First-generation speakers on both sides of the border come primarily from working-class backgrounds and their socioeconomic and educational levels are comparable to each other. It is the second-generation Tijuana speakers who overall exhibit a higher level of education. By having two groups (TJ1 and SD1) with a low-to-mid level of formal education in Spanish, a group with no formal schooling in Spanish (SD2), and a group with a mid-to-high level (TJ2), I am able to test for education as a factor that may lead to differences in grammar not only by comparing heritage speakers to non-heritage speakers, but also immigrant SD1 speakers to the Tijuana groups.

Chapter 3 Interactions and Exposure in Spanish and English

In chapter 1 I established that the following two aspects of the language experience of speakers in the border are the most salient for this study:

1. Interaction in Spanish and/or English
2. Exposure to Spanish and/or English

Interaction refers to the degree that each speaker interacts with others in Spanish or English or both languages. Exposure refers to reading, writing, watching television and movies, and listening to music in English and/or Spanish. The information about interaction and exposure for each speaker was gathered directly from the second part of the interviews.

One of the challenges in dealing with heritage populations is categorizing this group of speakers in such a way that their language ability can be clearly ascertained. By taking a close look at speakers' interactions and exposure to Spanish and English I can zero-in on the situations where speakers as a group show more variability in their Spanish/English language use. This approach underscores the importance of acknowledging that in a language contact situation – such as the one in the Tijuana-San Diego border area – language experience and use may be what shapes the linguistic system of speakers who have been removed from a non-contact variety of Spanish. In sections 3.1-3.3 I outline the most important aspects of the language experience of each group as they pertain to exposure and interaction in Spanish and English.

3.1 Tijuana: Interactions and Exposure

Speakers in Tijuana as a group represent a population that is not in significant contact with English so as to assume that their linguistic system is shaped by variable interaction and exposure to both English and Spanish. I provide only a brief overview for this group since their language experience and use is not characterized by language contact. The interactions for this group are mostly in Spanish, except for a few subjects who report using English occasionally when visiting the United States. For exposure, speakers in Tijuana report some alternation between English and Spanish, particularly watching movies and listening to music. In all, what characterizes the Tijuana language experience is that their interactions are Spanish-exclusive while their exposure to English is mostly passive and variable for some speakers and rare for the rest. There is no significant difference between the language experience and use of first and second generation speakers in Tijuana as it pertains to language contact.

3.2 Immigrants: Interactions and Exposure

The interactions that I present for first-generation speakers in San Diego are interactions with family, children, spouse, friends, workmates, and strangers. For exposure I present results for watching television and movies, and reading. The numbers in the tables represent the number of speakers (out of a total of 11) who report a particular aspect of language use in one of the following three categories: (1) Spanish only or mostly; (2) English only or mostly; or (3) both Spanish and English. Any cases

where the total does not add up to 11 speakers is due to speakers not reporting this aspect of language use in their interview. This information is summarized in figure 3.

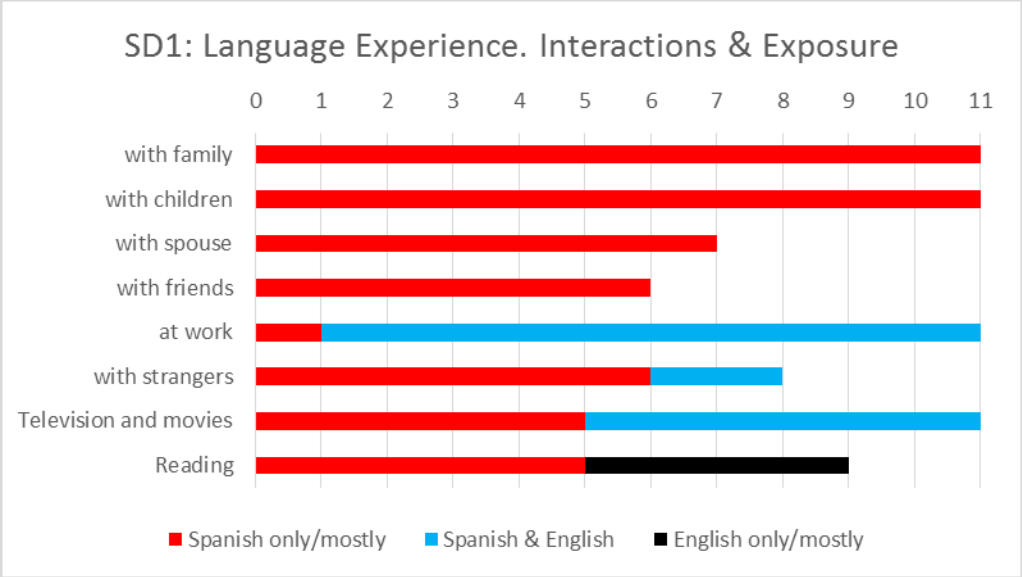

Figure 3.1 SD1 Interactions and Exposure.

Other than interacting with co-workers and with strangers, all of the reported interactions for the SD1 group are in only or mostly in Spanish. The only interaction that is not mostly or only in Spanish for almost all of the speakers (10 out of 11) is interaction with co-workers, for which SD1 speakers report interacting in both Spanish and English. With strangers, 75% (6 out of 8) of the interactions reported are in Spanish only or mostly. In terms of exposure, only reading is reported as an English only or mostly form of interaction for 4 out of 9 speakers. All four speakers report that the reading materials in English come from work, their children's school, their banks, or the government – in other words, official means of correspondence. None of the SD1 speakers report any significant amount of writing in any language. The picture that

emerges for interaction and exposure for the SD1 group is that their interactions are mostly in Spanish, except for work where they are variable Spanish with English. There is more variability in exposure between Spanish-mostly and Spanish/English for television and movies, but reading is done either exclusively in Spanish or in English. It is worth noting that the sole exposure to English-only for the group is reading, with 4 out of 9 speakers reporting it.

For formal education in English, 8 of the 11 speakers report having made attempts to learn English as a second language. For five for the speakers, the source of education in English was adult/night school, whereas for three of them it was through a combination of ESL school and at-home learning methods such as *Ingles sin Barreras*. Two of the speakers report not having had any schooling in ESL in the U.S. In Figure 3.2 I provide the breakdown of the source of education in English for the SD2 speakers.

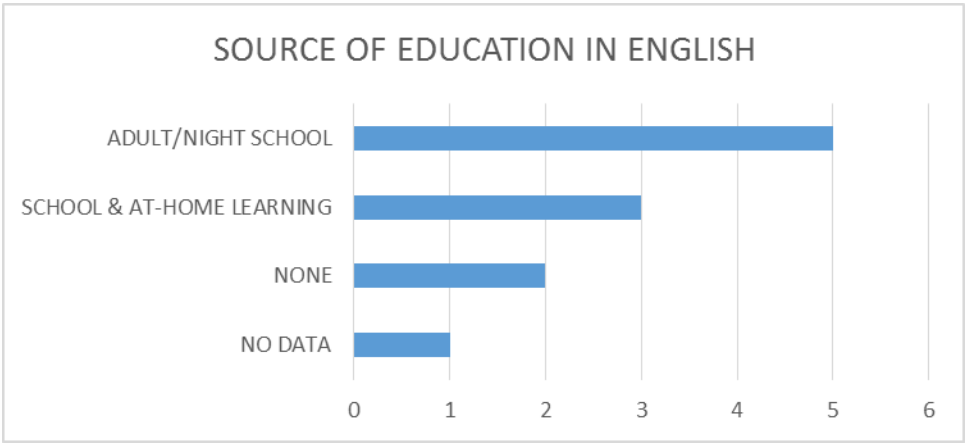

Figure 3.2 SD1 Source of Education in English in the U.S.

In regards to English as a second language courses taken in Mexico as part of the Mexican education system, 5 of the 11 SD1 speakers report having taken ESL courses in Mexico over 20 years prior to the time of the interview. Of these five, two speakers took these courses in *preparatoria* (high school) and three in *secundaria* (middle school). Four report having taken no English courses. This is presented in Figure 3.3 below.

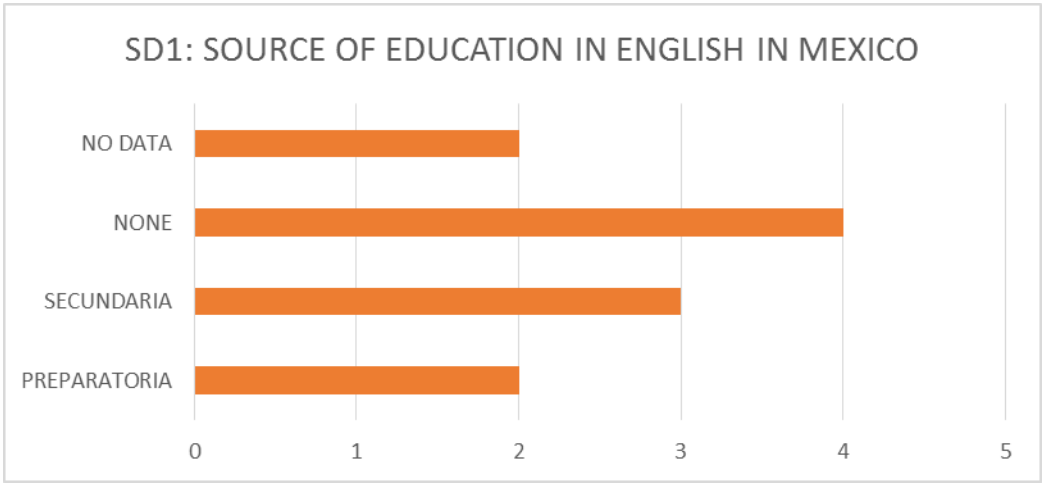

Figure 3.3 SD1 Source of Education in English in Mexico

When asked to report their level of proficiency in English, five speakers report speaking none to very little English; four report speaking it "more or less well" (*más o menos*); only two consider their English proficiency to be at a good level. These results are presented in Table 3.1 and they are self reported. No assessment of any of speaker's English-language skills was carried out during the interview for any of the populations.

Table 3.1 SD1 Self-Reported Proficiency Level of English

Level	Number of speakers
None to very low	5
More or less, okay (*más o menos*)	4
Good	2

In Table 3.2 I present a summary of the immigrant experience and its characteristics.

Table 3.2 The Immigrant Experience. Summary

Mean age	49.7
Years in Border	At least 18, up to 30
Socioeconomic status	Working class (9 out of 11, 81%)
Occupation	Working class/employee (8 out of 11, 73%)
Educational level	Wide range from finished *primaria* to 2 college graduates
Interactions primarily in Spanish	Family, friends, children, spouse, strangers
Interactions in Spanish/English	Work
Interactions primarily in English	None
Exposure primarily in Spanish	Television & movies, reading
Exposure in Spanish/English	Television & movies
Exposure primarily in English	Reading
ESL in US	7 out of 11 (64%) have some ESL
ESL in Mexico, prior to US	5 out of 11 (45.5%) took some ESL in Mex over 20 years prior to interview
Self-reported English	5 out of 11 (45.5%): none to very low 4 out of 11 (36.5%): okay 2 out of 11 (18%): good

3.3 Heritage Speakers: Interactions and Exposure

For second-generation speakers in San Diego the interactions that I report are language use with father, mother, with siblings, with other family members, with friends, and at work.. For exposure I present results for reading, writing, watching

television and movies, and listening to music. As with the first-generation group, this group reported their interactions and exposure to language as: (1) Spanish only or mostly; (2) English only or mostly; or (3) both Spanish and English. Any cases where the total does not add up to 11 speakers is due to speakers not reporting this aspect of language use. Figure 3.4 presents the interactions in and exposure to Spanish and English of heritage speakers.

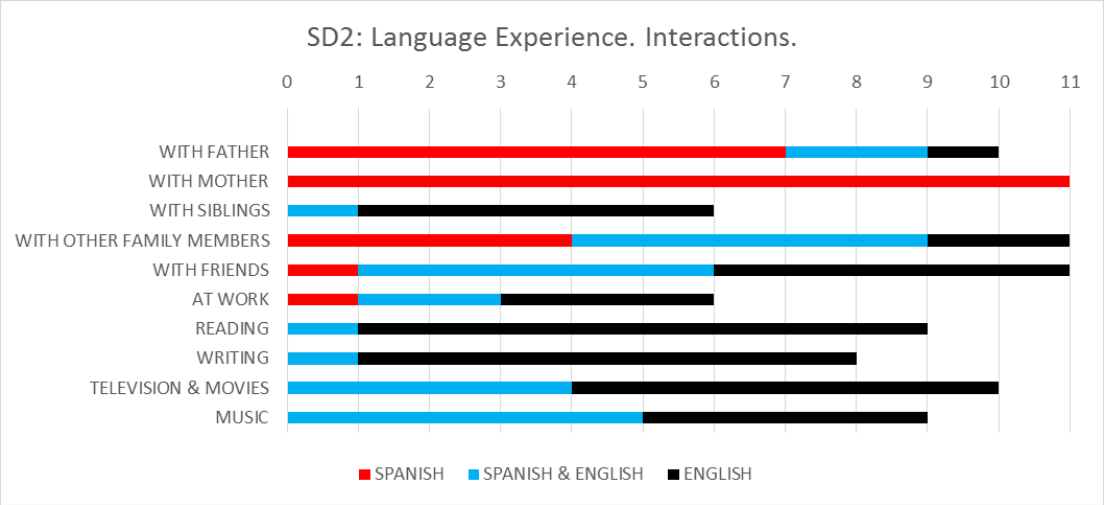

Figure 3.4 SD2 Interactions & Exposure

The only interaction that all speakers have exclusively in Spanish is with their mother. On the other hand, with their siblings the interactions are only in English for all but one speaker. The rest of their interactions (with father, with other family members, with friends, and at work) take place in both languages, either exclusively or in combination. The low number of speakers (5 out of 11) who report language use at work may be due to the fact that the majority of them are either recent high-school graduates

or first-year college students and therefore may not be employed. Other aspects of interaction that came up as part of the interview included language use at church by one speaker (Spanish only) and language use with a significant other (English-only for one speaker, and English and Spanish for another). The interactions for this group can be said to be defined by highly variable use of Spanish and English except with the mother (Spanish only) and siblings (English mostly).

In terms of exposure, heritage speakers do not report a single aspect of language use in Spanish only or mostly. Watching television and movies as well as listening to music occurs either in a combination of Spanish and English, or in English only. Reading and writing is done mostly in English for all speakers, except for one who reports English and Spanish for both reading and writing. Work by Valdes (1997), Potowski (2003), Potowski & Carreira (2004), and Montrul & Potowski (2007) among others has focused on pointing out the challenges that heritage speakers face in the heritage-language classroom as a result of poor reading and writing skills. Their work underscores the importance of reading and writing skills for the retention of Spanish in this population. As the data from the interviews in the corpus demonstrates, reading and writing in Spanish is scant for heritage speakers and the limited opportunities they have for literacy in Spanish are limited to what they receive in a Spanish classroom.

All 11 heritage speakers report Spanish as their first language and English as their second language. The acquisition of English coincides with formal schooling in English. For 3 of the 11 speakers this happens during kindergarten and for the rest during

the first four years of elementary school. In Table 3.5 I present the breakdown of the school grade that corresponds to the onset of English acquisition for all 11 speakers.

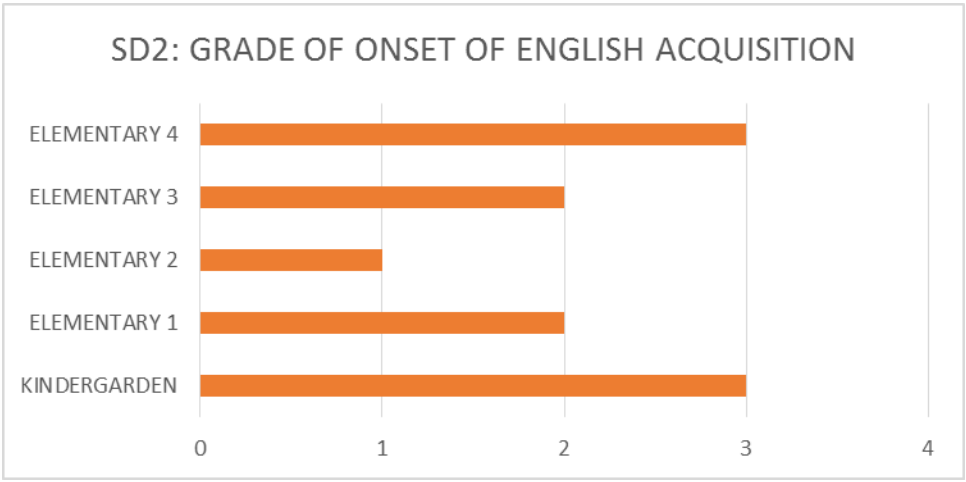

Figure 3.5 SD2 Onset of English Acquisition.

That the onset of English acquisition can span five different school grades indicates that the age at which a heritage speaker starts learning English can range from 4 years old to 8 years old. This process can also be gradual if the student is enrolled in a bilingual education program. Therefore, just as the language interactions and exposure of adult heritage speakers are characterized by great variability between Spanish and English, the process of becoming English dominant is also highly variable in terms of when it starts and how abrupt the shift is. Given that the interactions with siblings for many heritage speakers are English-only, it may be that the transition into English starts before the onset of English education for speakers who are not the first born. Unfortunately, information on order of birth was not collected for all speakers in the present study.

Potowski & Carreira (2004) stress the importance of exposure to Spanish through reading material in a Spanish-language classroom as one of the key factors that contributes to the retention of Spanish in heritage speakers. For speakers in the border area, this exposure may primarily occur in high school courses in Spanish. Unsurprisingly, in the present study the range in number of years of Spanish that students have taken in high school is wide. However, the majority of the heritage speakers (9 our of 11) report having taken one or two years of Spanish in high school. The breakdown of literacy in Spanish for all 11 speakers is given in Table F below.

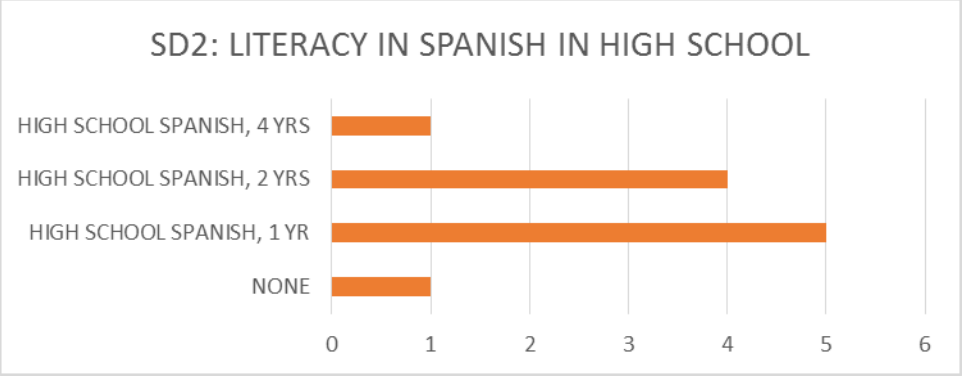

Figure 3.6 SD2 Literacy in Spanish in High School

It is worth pointing out that access to Spanish-language courses for heritage speakers vary from school to school. Some high schools may be able to offer four years of heritage Spanish whereas other schools may not be able to offer it at all. Literacy in Spanish (which affects interaction and exposure) is determined to some extent by the speaker's motivation and by availability of and access to heritage language courses.

In Table 3.3 I present a summary of the immigrant experience and what characterizes it.

Table 3.3 The Heritage Experience. Summary

Mean age	19.5 years
Socioeconomic status	Working class (9 out of 11, 81%)
Occupation	Students (9 out of 11, 81%)
Educational level	High-school or college (10 out of 11, 91%)
Interactions primarily in Spanish	Mother, Father
Interactions in Spanish/English	Friends, other family members, siblings
Interactions primarily in English	Siblings, friends, work, other family members
Exposure primarily in Spanish	
Exposure in Spanish/English	Television & movies, music
Exposure primarily in English	Reading, writing, television & movies, music
Onset of English acquisition	Wide range: Kindergarten to 4th grade
Literacy in HS Spanish	1-2 years of HS Spanish (9 out of 11, 81%)

3.4 Summary

The language experience of the San Diego populations can be said to be characterized by intense language contact and variability. San Diego speakers are in contact not only with English but also with Spanish in possibly two varieties: a contact variety that is a result of prolonged exposure to English, and a non-contact variety that is spoken by new arrivals and family members, and which may be present in the media as well. Immigrant speakers in San Diego vary greatly in their exposure and interactions in English, whereas heritage speakers exhibit a great range of variation in their exposure and interactions in Spanish.

By focusing on interaction and exposure as crucial components of language use and experience, one is able to take into account the great variability that speakers in a language contact situation exhibit. Moreover, among heritage populations in San Diego the degree and intensity of exposure and interactions in Spanish and English can shape the linguistic system of these speakers and result in a unique emerging system. The same may be true for the immigrant populations who arrive to the United States with a linguistic system in Spanish that may be similar to the Tijuana first-generation and after 20 or so years of intense contact (primarily two varieties of Spanish and to a variable extent English) may become its own system.

Chapter 4 Proficiency

One of the challenges of working with populations with varying degrees of literacy and formal education in Spanish is assessment. In chapter 3 I proposed that by focusing on interaction and exposure to Spanish and English among San Diego populations, one is better able to understand the great variability that characterizes Spanish in an intense language contact situation. In this chapter I present two methods of assessing the proficiency of Spanish speakers in the area. Section 4.1 focuses on lexical proficiency and speech rate as objective measurements of proficiency from the spoken data gathered in the interviews. Section 4.2 focuses on a subjective audio perception study wherein monolingual speakers in a different area of Mexico categorize a speech sample from each of the border speakers as belonging to one of the three language experiences: Tijuana, immigrant, and heritage. The goal of having objective and subjective assessments is to determine any differences between what speakers in the border produce (proficiency: objective assessment) versus how they are perceived (audio perception study: subjective assessment).

4.1 Objective Assessment: Lexical Proficiency and Speech Rate

Two important elements in assessing overall language fluency and proficiency are lexical proficiency and speech rate. Lexical proficiency consists of breadth and depth of knowledge of lexical items as well as access to these lexical items (Crossley et al 2011). Speech rate, which has been correlated to language proficiency (Polinsky 2008b), refers to the words per minute that a speaker produces. To determine the

difference in Spanish proficiency among the populations in the current study, I measure their lexical proficiency and speech rate to compare them with one another. The goal is to determine if immigrant speakers (SD1) differ in proficiency from their Tijuana counterparts (TJ1) and if heritage speakers (SD2) differ in proficiency from their parents (SD1) or from the Tijuana speakers (TJ1 and TJ2).

Other studies have observed the correlation between lexical proficiency & speech rate and overall language proficiency. Polinsky (1997, 2006, with Kagan 2007) has found a relationship between overall language proficiency and lexical knowledge from a 200-item word list when comparing heritage to native speakers in casual, conversational speech. Benmamoun, Montrul, and Polinsky (2010) also note that in comparison to native speakers, heritage speakers exhibit an attenuated knowledge of lexical items. The relationship between lexical and overall language proficiency has also been established in studies that focus on early childhood language development (Fenson et al 2000). For speech rate, Polinsky (2008b) found that among heritage speakers of Russian lower speech rate correlates to lower accuracy in gender restructuring. This same study by Polinksy (2008b) also notes that as heritage speakers have more difficulty accessing lexical items, their speech slows down and they resort to more pauses and fillers. Thus, lexical proficiency and speech rate appear to be related as measurements of overall language proficiency.

The proficiency of heritage language speakers has been assessed in other studies through cloze tests, fill-in-the-blank tests, and grammar tests. However, as Benmamoun, Montrul & Polinsky (2010) point out, these methods of testing can be problematic for

low-proficiency heritage speakers whose reading and writing skills may not allow them to complete the task. This is true for the speakers in the current study, not only the heritage group but also the SD1 and TJ1 groups. Some speakers move to Tijuana or the United States with a low-educational level and may have reservations about participating in a study that they may feel is testing their language ability The same applies to heritage speakers. Since the goal is first and foremost to obtain as much conversational and naturalistic data as possible, a pre-interview assessment may not put speakers at ease.

For this study, it was preferable to assess speakers' proficiency through conversational data as it presents a number of benefits. First, the interactional situation is the same for all four groups and no group of speakers was confronted with a topic that was outside of their language experience in a way that they may encounter in a cloze or fill-in-the-blank test. Given that heritage speakers may already feel insecure about their Spanish-language skills, feeling that they are being tested might exacerbate this negative self-perception and discourage them from participating.

Lexical knowledge and speech rate can be assessed in naturalistic conversational data. In a conversational setting the speakers draw from their lexical repertoire in context and in use. This is not something that can always be achieved through other methodologies such as a picture-naming task or a vocabulary test, which have been traditionally more appropriate for assessing L2 lexical proficiency. The conversational setting is also ideal for measuring speech rate in that it can be measured with familiar topics that the speaker chooses to discuss, as opposed to creating an artificial

environment through a description or a repetition task. In sum, considering the differences in formal education among the four groups, the variability in exposure and interactions in Spanish and English for the San Diego groups, and the English dominance of the heritage speakers, I forego more formal methods of assessment in favor of determining proficiency from conversational data.

Another benefit of drawing from conversational data is that it allows for speakers to be compared with one another as opposed to a canonical variety of Spanish, which is how more formal assessment methodologies operate. Speakers in San Diego are only compared to the counterparts in Tijuana. In terms of lexical proficiency, the following assumptions can be made about the populations:

1. TJ2 has the overall highest educational level in Spanish. This group is expected to have the highest lexical proficiency of all groups.
2. TJ1 and SD1 are comparable in education and socioeconomic status (refer to chapter 2). Any differences between these two groups, either in lexical proficiency and speech rate, may be due to the intense language contact situation that SD1 speakers find themselves in.
3. SD2 is expected to exhibit the lower lexical proficiency of all groups given their unique language experience that includes English dominance in interactions and lack of formal schooling in Spanish.

It is worth pointing out that the requirement for heritage speakers to participate in the study was the ability to carry out the interview in Spanish (even though codeswitching was not discouraged). This suggests that the group of heritage speakers

in this study is highly functional in Spanish. More than half of the SD2 participants were recruited by a high school teacher who may have knowledge of their Spanish-language ability. Hence, when I refer to heritage speakers in this study I refer to the 11 speakers in the SD2 group and not to all heritage speakers in the San Diego area[1].

The methodology and the results of the objective assessments are presented as follows: In section 4.1.1 I provide the methodology for measuring lexical proficiency, and in section 4.1.1.1 I present the results. Section 4.1.2 presents the methodology for measuring speech rate, and section 4.1.2.1.presents the results.

4.1.1 Methodology for Lexical Proficiency

The two measurements of lexical proficiency that I take into account for this study are noun and verb variety. Noun and verb variety refer to the unique number of nouns and verbs that speakers use. More proficient speakers are expected to employ a larger variety of unique nouns and verbs in Spanish whereas less proficient ones can be said to be operating under a more limited repertoire.

Verbs were tagged manually by the researchers. Both finite and non-finite verbs were taken into account. Nouns were tagged using an online part-of-speech tagger (textanalyzer.com) and then manually checked by the researcher for errors[2]. A unique

[1] The label "heritage speaker" can be problematic in that it is used to identify a wide spectrum of speakers with varying degrees of abilities and proficiencies in the heritage language. See Otheguy (2016) for a recent (and more complete) discussion of the shortcomings of the label "heritage speaker".

[2] For example, the part-of-speech tagger does not differentiate between *trabajo* ('work/I work') as a noun or as the first-person singular present-tense conjugation of *trabajar* ('to work'). Other nouns that required disambiguation from past-participle verbal forms include *puesto* ('a stand'/'put'), *hecho* ('a fact'/'done'), *dicho* ('a saying'/'said'), etc.

verb and noun ratio was then calculated for each speaker using the following formulas (1).

(1) Formula for unique verb and noun ratios

Unique verb ratio = number of unique verbs / total number of verbs used

Unique noun ratio = number of unique nouns / total number of nouns used

Not all nouns were used in the analysis. The following cases were omitted:

1. Proper nouns, numbers, and months & days of the week.
2. The nouns *español* ('Spanish') and *inglés* ('English') since the interview focuses mostly on language use and the repeated use of these two nouns would unbalance the ratios.
3. Acronyms.
4. Nouns in fillers such as *¿verdad?* ('right?').
5. Any English borrowings in San Diego not attested in Tijuana. For instance, 'sophomore' or 'prom'.

The nouns that were retained for analysis included English transfers in San Diego that have been lexicalized into Spanish and which have an equivalent in Tijuana with a different semantic interpretation (e.g., *grado* ('grade' for 'school year'), *papel* ('paper' for 'newspaper'), *parientes* ('relatives' for 'parents'), etc). Also retained were English borrowings that are attested on both sides of the border such as 'Spanglish', 'emails', 'mall', and 'swap meet'. All verb forms were considered as part of the ratio, including non-standard forms such as the subjunctive form *haiga* ('there is/are'). Omitted from

the analysis were incomplete verbs, verbs used as part of a discourse marker or a filler (such as *o sea* or *digamos*), and any verbs used in English as codeswitching.

A preliminary study revealed that taking into account the entire interview for each speaker resulted in unique noun and verb ratios that were not representative of the lexical repertoire of the speaker. Speakers with longer interviews were penalized with a lower unique noun and verb ratio since after a certain time they would start reusing previous nouns and verbs. To normalize the corpus, the number of nouns and verbs from the shortest interview (speaker SD-I1) was used as a benchmark. SD-I1 used a total of 247 verbs and 145 nouns. This was the number of verbs and nouns, respectively, that all interviews were reduced to. In all interviews, the last 247 verbs and 145 nouns were analyzed since these reflect the most conversational part of the interview. In (2) I present the updated formula for unique verb and noun ratios for the normalized corpus.

(2) Formula for unique verb and noun ratios

Unique verb ratio = number of unique verbs / 247

Unique noun ratio = number of unique nouns / 145

4.1.1.1 Lexical Proficiency in Tijuana and San Diego Speakers

I now present the results for lexical proficiency among speakers in the border area. First, for unique noun ratio there was no statistically significant difference among groups as determined by one-way ANOVA ($F(3,40) = 2.276$, $p = .094$). In Table 4.1 for each group I present the average number of unique nouns out of 145, the standard deviation, and the average unique noun ratio.

Table 4.1 Unique Noun Use

Group	Number of unique nouns / 145 (average for the group)	Standard Deviation	Unique noun ratio (average for the group)
TJ1	82.63	8.875	0.569
TJ2	81.81	8.177	0.564
SD1	79.81	7.493	0.550
SD2	74.27	6.889	0.512

In spite of the lack of statistical significance, the heritage speakers exhibit the lowest unique noun ratio followed by their parents. Tijuana second-generation speakers, surprisingly, do not emerge as the group with the overall highest unique noun ratio even though they are the group with the highest educational level in Spanish. The slightly reduced (though not statistically significant) noun repertoire for heritage speakers may be a result of a unique heritage language experience that is characterized by English dominance, reduced literacy in Spanish, and a variable and infrequent opportunities for interactions and exposure to Spanish (See chapter 2 on interactions and exposure).

For unique verb ratio, the difference between groups as determined by one-way ANOVA ($F(3,40) = 5.321$, $p = .00351$) is statistically significant. A post-hoc Tukey HSD test determined that the SD2 group differs significantly from the other groups. In Table 4.2 for each group I present the average number of unique verbs out of 247, the standard deviation, and the average unique verb ratio.

Table 4.2 Unique Verb Use

Group	Number of unique verbs / 247 (average for the group)	Standard Deviation	Unique verb ratio (average for the group)
TJ1	67.63	6.971	0.273
TJ2	73.09	8.184	0.295
SD1	68.63	7.327	0.277
SD2	60.36	6.314	0.244

It is with unique verb use that the predictions laid out at the beginning of this chapter are borne out. TJ2, with the overall highest educational level in Spanish, emerges as the group with the highest unique verb ratio though this is not corroborated by statistical significance. The TJ1 and SD1 groups, comparable in education and socioeconomic status, have an almost equal unique verb ratio. This may suggest that though these two groups may be similar sociodemographically, their different language experience does not translate into a significantly different repertoire of verbs. The lower lexical proficiency – but only as it pertains to verb use – for heritage speakers is corroborated by statistical significance.

The following aspects of the language experience of heritage speakers may bear upon the lower repertoire of verbs that they exhibit: (1) schooling in English, (2) reduced literacy in Spanish, and (3) English dominance in most interactions. Notice that contact with English in and of itself does not result in a reduced repertoire of verbs since no effect emerges for the SD1 group, who also has some contact with English. This finding supports Potowski & Carreira's (2004) point about the importance of literacy in Spanish, particularly reading, in retention of Spanish among heritage speakers. It may also be that verbs are more vulnerable than nouns to reduced formal schooling in

Spanish and to English dominance, as evidenced by the strong statistical significance for the heritage speakers.

4.1.2 Methodology for Speech Rate

To calculate the speech rate for each speaker, a two to three-minute segment of uninterrupted speech was extracted from each interview. The number of syllables in the segment was then determined. Each speaker's speech rate was calculated as in (3).

(3) Formula for speech rate

Speech rate = number of syllables/length of segment

4.1.2.1 Speech Rate in Tijuana and San Diego Speakers

For speech rate there was no statistically significant difference among groups as determined by one-way ANOVA ($F(3,40) = 0.734, p = .537$). In Table 4.3 for each group I present the average speech rate in words per minute and the standard deviation.

Table 4.3 Speech Rate

Group	Average speech rate (syllables per minute)	Standard Deviation
TJ1	138.50	26.110
TJ2	153.13	18.821
SD1	143.39	22.167
SD2	145.85	22.201

Though Polinski (2008b) notes that heritage speakers differ in speech rate when compared to the baseline speakers, this is not the case for the heritage speakers in San Diego that are part of this study. Moreover, they do not even represent the group with the lowest speech rate in the border area. This does not mean that speech rate is not an adequate measure of proficiency in heritage speakers. Rather, San Diego heritage speakers in this study may be past the proficiency threshold where there are any statistically-significant differences between them and the baseline (in this case all non-heritage speakers).

4.1.3 The Language Experience and Proficiency

Gender is the only external factor (out of all the sociodemographic and exposure/interaction outlined in chapter 3) that appears to have some relationship to lexical proficiency and speech rate. In this section I provide an analysis of how gender interacts with the different proficiency measurements but none of these results are corroborated by statistical significance. They merely point to trends that a larger population sample might reveal.

I begin with unique noun ratio. All groups except the Tijuana second-generation speakers (TJ2) exhibit a similar pattern: four out of five of the highest unique-noun ratios in each group belong a male speaker. These results are presented in Table 4.4 with the gender and the unique-noun ratio provided for the top five speakers per group. The unique-noun ratios are presented in descending order.

Table 4.4 Speakers with the highest unique-noun ratio (UNR) per group

TJ1			**TJ2**			**SD1**			**SD2**		
Spkr	**G**	**UNR**	**Spkr**	**G**	**UNR**	**Spkr**	**G**	**UNR**	**Spkr**	**G**	**UNR**
TJ-M1	M	0.676	TJ-B2	F	0.683	SD-K1	F	0.628	SD-M2	M	0.586
TJ-O1	M	0.648	TJ-M2	F	0.641	SD-J1	M	0.614	SD-I2	M	0.552
TJ-L1	M	0.614	TJ-S2	F	0.628	SD-M1	M	0.607	SD-C2	M	0.545
TJ-R1	F	0.593	TJ-O2	M	0.566	SD-L1	M	0.586	SD-J2	F	0.545
TJ-Q1	M	0.579	TJ-R2	M	0.559	SD-I1	M	0.559	SD-B2	M	0.531

Male speakers in TJ1 and in both generations in San Diego could exhibit a somewhat more diverse repertoire of nouns than female speakers. It is important to bear in mind that the interactional situation was completely autobiographical and speakers had the freedom to discuss any topics that pertained to their lives' histories and their language use. What this means for lexical proficiency is that speakers were not pushed outside of their comfort zone and the unique noun ratio refers to topics they chose to discuss.

For unique-verb ratio, the two noteworthy results pertain to second-generation speakers on both sides of the border. Four out of five of the highest unique-verb ratios among TJ2 speakers are female whereas for SD2 four out of five are male. These results are presented in Table 4.5 with the gender and the unique-verb ratio provided for the top five speakers per group. The unique-verb ratios are presented in descending order.

Table 4.5 Speakers with the highest unique-verb ratio (UVR) per group

TJ1			**TJ2**			**SD1**			**SD2**		
Spkr	**G**	**UVR**	**Spkr**	**G**	**UVR**	**Spkr**	**G**	**UVR**	**Spkr**	**G**	**UVR**
TJ-M1	M	0.344	TJ-B2	F	0.344	SD-C1	F	0.319	SD-I2	M	0.299
TJ-R1	F	0.291	TJ-Q2	F	0.344	SD-I1	M	0.307	SD-M2	M	0.283
TJ-U1	M	0.287	TJ-U2	F	0.336	SD-L1	M	0.299	SD-D2	F	0.259
TJ-O1	M	0.283	TJ-S2	F	0.307	SD-J1	M	0.295	SD-H2	M	0.242
TJ-T1	F	0.283	TJ-R2	M	0.295	SD-F1	F	0.295	SD-C2	M	0.238

Whereas men in all populations but TJ2 exhibit a higher unique-noun ratios than women, only SD2 males exhibit a higher unique-verb ratio over women. Women in TJ2 have a higher unique-verb ratio than men, and there appears to be no other relationship between gender and lexical proficiency.

The only relationship between speech rate and gender appears on first-generation female speakers on both sides of the border. For both TJ1 and SD1, four out of the five highest speech rates belong to women, as shown in Table 4.6 with the gender and the speech rate provided for the top five speakers per group. The speech rates are presented in descending order.

Table 4.6 Speakers with the highest speech rate per group

TJ1			**TJ2**			**SD1**			**SD2**		
Spkr	**G**	**SRate**	**Spkr**	**G**	**SRate**	**Spkr**	**G**	**SRate**	**Spkr**	**G**	**SRate**
TJ-P1	F	173.1	TJ-L2	M	184.6	SD-F1	F	173.2	SD-A2	F	192.3
TJ-L1	M	163.9	TJ-B2	F	178.0	SD-H1	M	171.0	SD-H2	M	168.8
TJ-S1	F	162.0	TJ-O2	M	171.5	SD-K1	F	163.8	SD-J2	F	161.6
TJ-F1	F	154.5	TJ-M2	F	169.8	SD-B1	F	154.1	SD-M2	M	155.7
TJ-T1	F	149.5	TJ-P2	F	150.7	SD-C1	F	150.4	SD-C2	M	146.2

Although there is no relationship between speech rate and proficiency, first-generation women emerge as the fastest speakers in the border area. Polinski (2008b) proposed that one of the likely reasons for a lower speech rate among heritage speakers is the preponderance of pauses and fillers. However, in the San Diego-Tijuana border area there is no correlation between the number of fillers a speaker uses and his/her speech rate. I provide a more thorough investigation of fillers in chapter 6.

4.1.4 Summary of the Objective Assessment of Proficiency

Lexical proficiency and speech rate have been identified as effective measurements of proficiency for heritage speakers. In the Tijuana-San Diego data for this study, only verb use appears to be an appropriate measurement to differentiate between populations. There is no statistically significant difference for noun use and speech rate among the populations. This may be due to the following:

1. Once the San Diego heritage speakers in this study attain a high level of competence, they cross a threshold of proficiency wherein their noun use and speech rate are indistinguishable for their parents and from the baseline speakers across the border.
2. One possible disadvantage of the autobiographical interview is that in asking all groups of speakers to talk about their lives, some of them might limit the breadth of topics they can discuss and thus not use their entire repertoire of nouns.
3. On the other hand, a possible advantage of the autobiographical interview is that in giving heritage speakers a chance to speak about topics that they are familiar with, and may even be used to speaking or thinking about (through interaction or exposure to the topic), they do not slow down their speech and therefore their speech rate is no different from any of the other groups.

Verb use emerges as the most accurate differentiator among populations. It is tied not only to educational level in Spanish but also to interaction and exposure to the language. SD2 speakers are statistically different from the other groups in their low unique-verb

ratio. What differentiates SD2 from the other population is English dominance and variability with Spanish in interactions and exposure as well as a lack of formal education in Spanish. Even in an interactional situation that is within the comfort zone of heritage speakers in San Diego-Tijuana, the results from the proficiency measurements for this study suggest that verb use is the most accurate indicator of proficiency differences among heritage and non-heritage populations. This may suggest that verbs, and not nouns, are more vulnerable to English dominance and lack of schooling in Spanish.

4.2 Subjective Assessment: Audio Perception Study

Benmammoun, Montrul, and Polinsky (2008) note anecdotally that native speakers are able to identify non-native speakers in under a minute or so. Heritage speakers, however, complicate this issue since their competence level is highly variable. Other studies (Major 1992; Schmid 2009; de Leeuw et al. 2010) suggest that immigrant speakers may over time be perceived as non-native due to lack of exposure and possibly interaction in their native language.

Considering the proximity of San Diego to Tijuana as well as the unique language contact situation in San Diego, is there any evidence for an immigrant accent or a heritage accent that non-immigrant monolingual speakers in Mexico detect? An audio perception study was carried out to investigate if the populations under study were perceived by Mexican monolinguals as sounding different from one another. The goals of this audio perception study are the following:

1. To determine if a population not in contact with the various varieties of Spanish in San Diego-Tijuana border be able to tell them apart.
2. To determine if heritage speakers are distinguishable from immigrant and Mexican speakers.
3. To determine if a distinct accent develops as a result of immigration in the San Diego first-generation speakers.
4. To determine if there is a border accent that may be distinct from interior Mexican accents.

4.2.1 Methodology

Eight monolingual Mexican speakers of Spanish in the city of Chihuahua, Mexico, (sometimes referred to as "listeners" in this chapter) were asked to categorize each of the 44 speakers in the corpus as belonging to one of the following three categories:

1. *Mexicano*: someone born and raised in Mexico and living somewhere in Mexico.
2. *Inmigrante*: someone born and raised in Mexico but currently living in the United States, who immigrated to the United Sates as an adult and may or may not speak English.
3. *Americano*: someone born and raised in the U.S. to Spanish-speaking parents, who speaks Spanish at home but English elsewhere and whose first language is Spanish.

Each audio clip for each speaker lasted between 30-60 seconds. The segment from the interview that was chosen did not include any code switches, filled pauses in English, and no information hinting at where the speaker may be living or may be from. The Chihuahua listeners were not given any information regarding the speakers heard on the clips other than the fact that each speaker belonged to one of three categories given.

None of the Chihuahuan listeners have ever lived in the United States or in the U.S.-Mexico border area, though they all report having visited the U.S. as tourists and having family members in the United States. Therefore, they are all familiar with both immigrant and heritage speakers of Spanish in the United States.

4.2.2 Results and Discussion

I now present the results of the audio perception study for each of the four groups. For each speaker in TJ1, TJ2, SD1 and SD2 I count the number of times that he or she was identified as *mexicano*, *inmigrante*, or *americano* by each of the eight Chihuahua listeners. I then calculate the average number of IDs (or judgments) of *mexicano*, *inmigrante*, or *americano* for each group (TJ1, TJ2, SD1 and SD2). For example, table 4.7 shows that speakers in TJ1 as a group have an average number of IDs of *mexicano* of 3.82 out of a possible 8 (since there are eight listeners). The combined ID averages of *mexicano* and *inmigrante* correspond to the non-heritage category, whereas the combined ID averages of *inmigrante* and *americano* correspond to the US speaker category. Table 4.7 contains the results of the audio perception study for the Tijuana and San Diego groups.

Table 4.7 Audio Perception Study

Perceived by Chihuahua listeners…	Average N of IDs (out of 8 listeners)			
	for TJ1	for TJ2	for SD1	for SD2
… as *mexicano* (TJ)	3.82	3.09	1.73	1.00
… as *inmigrante* (SD1)	3	3.73	5.45	3.55
… as *americano* (SD2)	1.18	1.18	0.82	3.45
… as non-heritage (TJ & SD1)	6.82	6.82	7.18	4.55
… as a US Speaker (SD)	4.18	4.91	6.27	7.00

The results for both Tijuana groups are similar in that the averages for both groups in the three discrete categories (*mexicano*, *inmigrante*, and *americano*) and the two combined categories (non-heritage and US speaker) fall within 1.00 ID of each other. In fact, the averages are the same for the two groups for the categories *americano* and non-heritage. Both Tijuana groups are more accurately identified as non-heritage (6.82 average for both) than as US speakers (4.18 for TJ1, 4.91 for TJ2). For both Tijuana groups, the split between the categories of *mexicano* and *inmigrante* may be due to the fact that Chihuahua listeners perceive Tijuana speakers as sounding different from themselves. Since Chihuahua listeners identify as *mexicano*, the category *inmigrante* represents the only "other" category that is neither *mexicano* nor *americano*. As such, these results do not necessarily suggest that Tijuana speakers are perceived as immigrants to the United States. Rather, they may be perceived as different from Chihuahua speakers using the only "other" category that Chihuahua listeners have at their disposal. I conclude that both generations of Tijuana speakers are not perceived as significantly different from each other.

San Diego first-generation speakers are perceived as a group separate from the Tijuana speakers as evidenced by the differences in averages between TJ and SD1 in

the discrete categories *mexicano* and *inmigrante* and the combined category US Speaker. The high average of 5.45 *inmigrante* IDs and the low average of 1.73 *mexicano* IDs for the SD1 group indicates that Chihuahua speakers perceive SD1 speakers differently from Tijuana speakers. This indicates that there may be phonetic details such as intonation or vowel quality that Mexican immigrants to the United States acquire after having lived over 18 years in the U.S. In chapter 2 on demographics and populations I noted that the 11 SD1 speakers were originally from six different states: Jalisco, Sinaloa, Baja California Norte, Chihuahua, Queretaro, and Oaxaca. It is not surprising that after 18 years of being removed from their original varieties of Spanish and being in intense contact with the San Diego variety, the SD1 speakers may develop phonetic properties that are perceptively different from *mexicano* and *americano* speakers. Lastly, another significant difference between Tijuana and SD1 speakers is the average number of IDs for US Speaker: at 6.27 for the SD1 speakers it is significantly higher than the averages of 4.18 and 4.91 for TJ1 and TJ2 respectively.

Second-generation heritage speakers in San Diego (SD2) are perceived differently from TJ and SD1 speakers in the discrete categories *mexicano* and *americano* and in the combined category non-heritage. SD2 speakers are judged as being the least *mexicano* (ID average 1.00) and the most *americano* group (ID average 3.45) in the border area. They are also judged as being the least non-heritage (4.55 average ID) and the most US Speaker (7.00). These perceptible differences among SD2 speakers may be the result of a unique set of phonetic properties that differentiates SD2 speakers from both SD1 and TJ speakers.

The most unfortunate limitation of the audio perception study is the number of listeners. A larger listener sample might yield stronger results. Another limitation is the number of clips that listeners had to judge. Some reported being bored by the 44 clips and therefore their concentration and accuracy might decrease as the task progresses[3]. Finally, carrying out the same task with different populations of listeners (particularly listeners from the same populations being tested) may reveal how accurate speakers are at identifying their own variety of Spanish.

4.2.3 Summary of the Subjective Assessment: Audio Perception Study

To return to the questions laid out at the beginning of this section, it would appear that the group of native monolingual listeners in Chihuahua perceive three distinct categories of speakers of Spanish in the border area:

1. A Tijuana category that consists of TJ1 and TJ2 speakers.
2. An immigrant category for SD1 speakers. Although SD1 speakers are Spanish-dominant, prolonged contact (18+ years) with the heritage variety of Spanish and to a lesser extent English – as well as reduced exposure to the non-contact varieties of Mexican Spanish – may lead to phonetic properties that make SD1 speakers perceptively different from Tijuana speakers but also dissimilar from SD2 speakers.

[3] The audio clips were played in random order to each listener. Each listener heard the audio clips in a unique sequence.

3. A heritage category for SD2 speakers. English-dominance for the SD2 speakers may be the cause for the presence of phonetic features that perceptively set apart heritage speakers from their parents and from Tijuana speakers. What reinforces the notion of the SD2 speakers in the present study as a highly-functioning and proficient group of speakers of Spanish is that the average number of IDs in the category *inmigrante* (3.55) is slightly higher than the average for *americano* (3.45). This suggests that some heritage speakers, in spite of their English dominance, may at times be perceptively indistinguishable from their Spanish dominant parents.

4.3 Conclusion

The objective assessment of proficiency reveals that the only statistically-significant difference for any group is verb use for heritage speakers. Heritage speakers exhibit a lower unique-verb ratio that is statistically different from the other populations. This may point to a more limited verb repertoire for heritage speakers that is a result of English dominance and variability with Spanish in interactions and exposure to the language, as well as lack of formal schooling in Spanish. In terms of noun use and speech rate, no significant difference was found among any of the groups. The subjective assessment indicates that three distinct groups of Spanish speakers can be perceived in Tijuana and San Diego by a group of listeners from outside the border area. These groups are: Tijuana, immigrant, and heritage. Interactions and exposure in Spanish and English, as well as language dominance in English, are likely the factors

that lead to the emergence of phonetic features that distinguish the two groups of speakers in contact with English (SD1 and SD2) from each other and from speakers in Tijuana.

In spite of being perceived somewhat differently from non-heritage speakers by monolingual speakers in Mexico, the only difference in proficiency between heritage and non-heritage speakers is unique verb use in that heritage speakers as a group employ less unique verbs in their interviews. There are no differences in noun use or speech rate among the groups. The SD2 speakers in this study are highly proficient and highly functional in Spanish, and they do not represent a group that is radically different from their parents or comparable populations in Tijuana in terms of proficiency or functionality in Spanish.

Chapter 5 The Subjunctive in the Border Area

In this chapter I focus on mood selection in the Tijuana-San Diego border area. The subjunctive has been identified as a particularly vulnerable area not only in contact varieties of Spanish but also in monolingual ones such as those found in Argentina (Manegotto 2005), Spain (Gallego & Marks 2014), the Canary Islands (Serrano Montesinos 1992), and Costa Rica (Kowal 2007 cited in Gallego & Marks 2014) to name a few. In fact, simplification of the subjunctive is a phenomenon that is happening diachronically within Romance languages (Silva-Corvalán 1994, Quer 2001) such as in Portuguese (Silva 2008), French (Borgonovo, De Garavito, Prevost 2008), Catalan and Italian (Quer 2010). Silva-Corvalán (1994) proposes that while the simplification of the mood system is ongoing in monolingual varieties of Spanish, a language contact situation may accelerate the simplification or loss of features of the tense-mood-aspect system. Of particular interest to this study are the effects that English dominance and lack of formal schooling in Spanish may have on the San Diego-born speakers' use of subjunctive.

5.1 The Subjunctive

In most cases, the subjunctive mood in Spanish occurs in embedded clauses where the matrix clause has created an environment that selects for subjunctive use. Mood selection can be said to be obligatory or optional; in these latter contexts the subjunctive mood can alternate with the indicative.

In an obligatory context mood alternation is not possible in the embedded clause and only subjunctive is licensed. In this study I refer to obligatory contexts as S-ONLY contexts. Examples (1-4) constitute S-ONLY contexts extracted from the Tijuana-San Diego data from the present study. For each of the examples I provide an alternate in the indicative mood (*) to show its ungrammaticality.

(1) a. *Yo he tratado de empujar a que sigan adelante* [SD-J1]

b. *Yo he tratado de empujar a que *siguen adelante*

'I have tried to push them to keep forging ahead'

(2) a. *No me gusta. Quiero que sepan los dos (idiomas)* [SD-C2]

b. *No me gusta. Quiero que/*saben los dos (idiomas)*

' I don't like it. I want them to know both (languages)'

(3) a. *Estamos a punto de que nos entreguen nuestra casa* [TJ-T1]

b. *Estamos a punto de que nos *entregan nuestra casa*

'We are about to receive our house'

(4) a. *El problema que yo espero que sí se solucione* [TJ-L2]

b. *El problema que yo espero que sí se *soluciona*

' The problem that I hope will get solved'

Obligatory contexts are attested in all four populations in the border area. Also included for analysis as S-ONLY contexts are expressions of volition or wishes where the main clause has been omitted. These expressions can be interpreted contextually as volitional or wishes in response to a question. Examples (5-8) from the corpus belong in this category.

(5) *Que terminen una carrera y que sean felices* [SD-E1]

'(I hope) that they complete a degree and be happy'

(6) *Que no tengan vergüenza de hablar su idioma* [SD-L2]

'(I want) for them to not be ashamed of speaking their language'

(7) *La educación, que se fomentara más* [TJ-M1]

'(I would like) for education to be promoted more'

(8) *Los lugares que están muy abandonados, que los ocuparan para algo más productivo* [TJ-P2]

'The places that are abandoned, (I wish) that they would be used for something more productive'

In alternating contexts the subjunctive alternates with the indicative in the embedded clause. I refer to these contexts as S-ALT in this study. Examples (9-12) present S-ALT contexts from the corpus. Below each of the examples I provide a variant in the alternate mood to illustrate the grammaticality of both selections of mood.

(9) a. Ind: *Quizás son 95% pero prefiero hablar español* [SD-C1]

b. Subj: *Quizás sean 95% pero prefiero hablar español*

'Maybe it is/could be 95% but I prefer to speak Spanish'

(10) a. Ind: *Me desespera que la gente es cochina* [TJ-S2]

b. Subj: *Me desespera que la gente sea cochina*

'It irritates me that people are/could be such pigs'

(11) a. Subj: *No hablar otro idioma que otra gente no pueda hablar* [SD-B2]

'To not speak another language that other people may not be able to speak'

b. Ind: *No hablar otro idioma que otra gente no puede hablar*

'To not speak another language that other people cannot speak'

(12) a. Subj: *Los que tengan billetes son los que van a estudiar* [SD-J1]

'Those who might have money are the ones who are going to study'

b. Ind: *Los que tienen billetes son los que van a estudiar*

'Those who have money are the ones who are going to study'

All four populations in the border area produce alternating contexts.

There are four morphosyntactic categories for the subjunctive in standard Spanish: present subjunctive (*hable*), present-perfect subjunctive (*haya hablado*), imperfect subjunctive (*hablara*), and pluperfect subjunctive (*hubiera hablado*). These four categories are theoretically possible in both S-ONLY and S-ALT contexts in Spanish. However, present perfect and pluperfect subjunctive are relatively rare in the border area across all populations. In the literature these morphosyntactic forms are sometimes identified as tense. Though I do use the label TENSE to refer to these forms as a variable in the logistic regression analysis of the subjunctive, in the study I refer to them as morphosyntactic categories or forms.

The subjunctive is also attested in non-subjunctive contexts where the conditional is normally expected, for example in the apodosis clause of a hypothetical sentence (examples 13-14), in an assertion in the main clause (15), and in an embedded

clause where the subjunctive mood is not normally selected (16). In all the examples below I provide the conditional alternative that might be expected in these contexts.

(13) a. Subj: *Si estuviera viviendo en Mexicali pues tuviera un trabajo* [SD-H2]

b. Cond: *Si estuviera viviendo en Mexicali pues tendría un trabajo*

'If I were living in Mexicali, I would have a job'

(14) a. Subj: *Tal vez si yo fuera a trabajar allá pues fuera diferente* [TJ-V1]

b. Cond: *Tal vez si yo fuera a trabajar allá pues sería diferente*

'Maybe if I went there for work, it would be different'

(15) a. Subj: *Sí estuviera diferente pero no sé cómo* [SD-B1]

b. Cond: *Sí estaría diferente pero no sé cómo*

'Yes, it would be different but I don't know how'

(16) a. Subj: *Siento que estuviera, no sé, más aburrida mi vida* [TJ-T2]

b. Cond: *Siento que estaría, no sé, más aburrida mi vida*

'I feel that my life would be, I don't know, more boring'

I refer to these contexts as COND-EXP and the only morphosyntactic categories in which they are attested are the imperfect and pluperfect subjunctive. All populations in Tijuana and San Diego produce these types of non-subjunctive contexts. Moreover, this use of the imperfect and pluperfect subjunctive has been attested among U.S. speakers of Spanish of Mexican descent in Los Angeles (Silva-Corvalán 1994), of Puerto Rican descent in New York (Torres 1997), and of Puerto Rican, Dominican, Mexican, Ecuadorian, Colombian, and Cuban descent in New York City (Bookhamer 2013).

Excluded from consideration in this study are uses of the subjunctive in discourse markers or fillers (e.g., *o sea*, *digamos que*, *supongamos que, ¿cómo le dijera?*) where speakers use the subjunctive in a fixed expression where no mood selection is required. Keeping with the focus on subjunctive vs. indicative, also excluded are forms of the imperative, both formal and informal.

5.2. Previous work on the Subjunctive in U.S varieties of Spanish

In this section I provide a brief description and summary of the major findings of work whose data, approach, and methodology are similar to this study: data that is gathered through naturalistic oral production in interviews, where more than one generation or population is taken into account, and where at least one of these populations consists of heritage speakers of Spanish[1].

One of the major studies on Spanish in the United States is Silva-Corvalán's work on Spanish in Los Angeles among Mexican immigrant and heritage speakers (Silva Corvalán 1994, 1995, 2001). Her work focuses on the tense-mood-aspect systems of three generations of speakers (first, second, and third generations) and the external and internal factors that motivate language change (1994). With respect to the mood system, Silva-Corvalán reports that a pattern of reduction of subjunctive use emerges with every subsequent generation. S-ALT contexts are particularly vulnerable to simplification (defined in her study as the extension of indicative over subjunctive) since

[1] I am not considering experimental work on the Spanish subjunctive since the methodology varies significantly from that of the present study.

using the indicative in these contexts is not ungrammatical. Beginning with the second generation, heritage speakers undergo simplification of the subjunctive in S-ALT contexts and of the present perfect and pluperfect subjunctive forms. Heritage speakers also exhibit an increased use of the indicative (particularly the conditional) in hypothetical conditionals in comparison to first-generation speakers.

Silva-Corvalán (1994) concludes that the simplification of the subjunctive among heritage speakers does not serve as evidence of direct influence from English since heritage speakers retain the subjunctive in S-ALT contexts. The simplification of the subjunctive may be due instead to reduced exposure to Spanish. First-generation speakers may also gradually exhibit simplification of the subjunctive in variable contexts after prolonged exposure to a variety of Spanish that has already undergone simplification and may have reduced normative standards (Silva-Corvalán 2001).

Also looking at three generations of speakers of Spanish in the U.S. is Lynch (1999), whose work on Spanish among Cuban Americans in Miami focuses on the subjunctive mood. The subjunctive mood is in use in all three generations of Cuban Americans but the system shows more variability in the second and third generations. The obligatory contexts for subjunctive appear to be stable for the most part and display no significant difference in variation across all three generations of Cuban Americans. The greater statistical effect for simplification of the subjunctive in variable contexts occurs in the third generation though some changes in the mood system are attested in the second generation as well. The imperfect and pluperfect subjunctive are also used in hypothetical conditionals in place of the conditional particularly by the second and

third generations. Lynch (1999) observes that the changes in the mood system among Cuban Americans in Miami align with those of Silva Corvalán (1994) for Mexican-American speakers in Los Angeles but at a much slower rate since the strongest effect in Miami is on the third generation.

Montrul (2001) analyzes oral and written data from U.S.-born Mexican-American heritage speakers with a language experience similar to the heritage speakers in San Diego in the present study. Montrul divides the heritage speakers according to proficiency in Spanish and compares them to native-speaker controls. In the oral task, high-proficiency speakers do not perform significantly different from the first-generation controls. Mid and low-proficiency speakers perform significantly differently from the controls in that their subjunctive use is simplified in S-ALT contexts. In the written and grammaticality judgement tasks, all three heritage speaker groups perform at a level below the native controls. Montrul (2009) ascribes these differences to a combination of attrition and incomplete acquisition wherein mood selection, given its complexity, is not fully acquired during childhood language development and as such its features remain vulnerable and they are the first to be lost.

Bookhamer's (2013) work on first and second generation speakers in New York City draws from the Zentella-Otheguy corpus (Otheguy & Zentella 2012), a corpus that represents a diverse group of speakers of Puerto Rican, Dominican, Mexican, Ecuadorian, Colombian, and Cuban descent. Heritage speakers across the board exhibit a statistically-significant lower frequency of subjunctive use when compared to the first generation. However, there is no difference in heritage speakers' command of

subjunctive morphology. Bookhamer proposes a "fragmented mood grammar" for heritage speakers that is characterized by great variability in pragmatic use but not in morphosyntactic production of the subjunctive.

5.3 Research Questions and Methodology

Based on previous studies, the following research questions guide this investigation of mood selection and the morphosyntactic categories of the subjunctive in the Tijuana-San Diego border area.

1. What is the frequency of the subjunctive in the border area and how does it compare to other varieties of Spanish in the U.S.? Do any external sociodemographic factors play a significant role in any difference in subjunctive use?
2. Are there any statistically significant differences in mood selection in S-ONLY and S-ALT contexts among the four populations?
3. Are the morphosyntactic categories of the subjunctive stable for all four populations or do any of these categories undergo simplification or attenuation?
4. Given that the subjunctive mood is extended to contexts where the conditional is expected, is this change systematic and does it apply to all four populations in the border differently?

The present study contributes to a growing body of research on the mood system of different varieties of Spanish in the U.S. since it is the first study of subjunctive use

in the San Diego area. In chapter 4 I established that second-generation San Diego speakers are highly proficient and functional in Spanish with the only difference between them and the non-heritage speakers being unique verb use. Two particular areas where heritage speakers differ from non-heritage speakers in language experience are language dominance and access to formal schooling in Spanish. Given that heritage speakers are highly proficient and functional in Spanish, do they use the subjunctive differently from non-heritage speakers given that they are English dominant and that they have had no formal schooling in Spanish?

In chapter 2 I presented the methodology for gathering and coding the data, and provided a detailed description of the nature of the interviews. In the same chapter I also outlined the sociodemographic factors that describe each population and their language exposure and interactions. To analyze mood selection and the morphosyntactic categories of the subjunctive, I consider only the verbs that occur in S-ONLY, S-ALT, and COND-EXP contexts and their morphosyntactic categories. For each group I run a Shapiro-Wilk Normality Test to determine if the population can be assumed to demonstrate normal distribution. The null hypothesis in the Shapiro-Wilk Normality Test is that the data is normally distributed. Therefore, a p-value lower than 0.01 would indicate that the data is not normally distributed. I report the results of this test for normality in terms of a p-value. If the data appears to have normal distributions, I employ the following two tests to compare the groups:

1. A one-way analysis of variance (ANOVA) then tests whether any of the four populations are statistically different from one another. In such case, a post-

hoc Tukey HSD test determines which group is different from the rest, or if more than one group behaves differently.

2. A two-tailed T-Test to compare heritage to non-heritage speakers to determine if there is a heritage effect, and San Diego immigrants to Tijuana speakers to determine if there is an immigrant effect.

Lastly, a logistic regression analysis using Rbrul allows for the analysis for mood selection and the morphosyntactic categories of the subjunctive incorporating sociodemographic variables as external predictors; mood, context, and tense as internal predictors; individual speakers as random effects in case the data does not pass the Shapiro-Wilk Normality Test due to it not being normally distributed. It is worth mentioning that a lack of a statistically-significant effect for any of the comparisons is just as informative as the presence of a statistically-significant effect since it would reveal that in spite of the different language experiences for the groups (particularly in regards to language dominance and access to formal schooling in Spanish), some aspects of the subjunctive system may not be affected. I present the results in the following section.

5.4 A Quantitative Analysis of Mood Selection and the Morphosyntactic Categories of the Subjunctive

My corpus contains a total of 20,481 finite verbs, of which 1086 are in the subjunctive mood. This means that 5.3% of the finite verbs produced are in the subjunctive. Subjunctive use in the Tijuana - San Diego area is slightly less frequent

than in Otheguy-Zentella's New York City corpus where immigrant speakers produce 6.8% of finite verbs in subjunctive and heritage speakers 5.2% (Bookhamer 2013). To determine if there is any statistically significant difference in subjunctive frequency among the populations, a percentage of subjunctive frequency was calculated for each speaker by taking into account each speaker's number of finite verbs and the number of subjunctive tokens. In Table 5.1 I present the percentage of subjunctive use for all speakers, organized by group and sorted in descending order from largest to smallest, and the average percentage of subjunctive use for each population.

Table 5.1 Frequency of Subjunctive use in the Border Area

Group	TJ1	TJ2	SD1	SD2
Percentage of subjunctive use for each speaker per group = speaker's number of subjunctive tokens / speaker's number of finite verbs	8.28 7.33 6.91 5.56 5.16 5.11 4.32 3.17 3.07 2.61 1.93	10.18 7.73 6.99 6.77 6.08 5.90 5.76 5.71 5.56 3.45 1.60	9.64 6.36 5.92 5.74 5.60 4.60 4.18 3.75 2.82 2.01 1.64	8.62 6.66 6.45 6.45 5.07 4.90 4.70 3.79 2.69 2.53 1.16
Average for group	4.86	5.98	4.75	4.82
Standard deviation	1.96	2.09	2.17	2.07
Shapiro Wilk Normality Test p-value	p>0.05	p>0.05	p>0.05	p>0.05
Data normally distributed?	yes	yes	yes	yes

Speaking in broad terms, frequency of subjunctive use is the highest for the TJ2 population and the lowest for the SD1 population. However, a one-way ANOVA ($F(3,40) = 0.790$, $p = .506$) shows that there is no statistically significant difference in

frequency of subjunctive use among the populations. No statistically significant difference is revealed by two-tailed T-Tests comparing non-heritage speakers to heritage speakers ($p<0.622$) or immigrant to Tijuana speakers ($p<0.417$). Another aspect worth noting is the wide range of variation within populations. Even the TJ2 group, with the overall highest level of formal educational in Spanish, exhibit this wide range of variation in frequency of subjunctive. One of the advantages of working with naturalistic conversational data is that it can reveal the great range of variation that exists in subjunctive use within populations regardless of their level of proficiency.

In the sections that follow I present the ratios of subjunctive use for the three contexts where the subjunctive mood is attested in the study: S-ONLY, S-ALT, and COND-EXP. For each of these I provide a breakdown of the morphosyntactic categories for the four populations and whether or not there is any statistically significant difference among groups for each of the contexts and their corresponding morphosyntactic categories.

5.4.1 The Subjunctive in Obligatory Contexts

Taking into account the entire number of finite verbs for each speaker, I ask the following: what percentage of all finite verbs corresponds to verbs used in an S-ONLY context? By comparing the percentages for all four populations, I can determine if any group(s) create more or fewer S-ONLY contexts than any other group(s) in a manner that is statistically significant. In Table 5.2 I present the percentages of tokens used in an S-ONLY context sorted in descending order.

Table 5.2 Percentage of S-ONLY Contexts

Group	TJ1	TJ2	SD1	SD2
Percentage of S-ONLY contexts = speaker's number of tokens in S-ONLY contexts / speaker's number of finite verbs	6.29 5.20 3.17 3.00 2.99 2.80 2.78 2.40 2.33 2.31 1.93	6.94 5.41 5.08 5.00 4.05 3.79 2.71 2.43 2.38 2.02 0.53	6.76 5.53 4.55 3.83 3.38 3.08 3.07 1.99 1.65 1.51 1.41	5.58 5.00 3.49 3.43 3.38 3.23 3.20 3.13 1.80 0.84 0.70
Average for group	3.19%	3.66%	3.34%	3.07%
Standard Deviation	1.26	1.76	1.65	1.436
Shapiro Wilk Normality Test p-value	$p<0.01$	$p>0.05$	$p>0.05$	$p>0.05$
Data normally distributed?	no	yes	yes	yes

Though the range of variation is wide within each of the groups and the averages for each group are not very different from one another, the distribution of the data for the TJ1 population. This means that statistically significant difference cannot be determined with a parametric test. A logistic regression analysis in section 5.5 may help determine if the populations are significantly different in S-ONLY contexts.

Though S-ONLY contexts only license the subjunctive in the embedded clause, some speakers may use indicative or conditional. Taking into account only S-ONLY contexts, in Table 5.3 I provide the percentage of S-ONLY contexts with a subjunctive verb in the embedded clause. The percentages are presented in descending order along with an average for each of the groups. Notice that for any speakers where their percentage is not 100 (four speakers in the TJ2 and SD2 groups, and one in the SD1 group), it indicates use of the indicative mood in S-ONLY contexts.

Table 5.3 Percentage of Subjunctive Verbs in S-ONLY Contexts

Group	TJ1	TJ2	SD1	SD2
Percentage of subjunctive verbs in S-ONLY contexts = speaker's number of subjunctive tokens in S-ONLY contexts / speaker's number of overall tokens in S-ONLY contexts	100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00	100.00 100.00 100.00 100.00 100.00 100.00 100.00 96.67 95.24 95.24 92.86	100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 98.78	100.00 100.00 100.00 100.00 100.00 100.00 100.00 88.89 87.50 85.71 70.00
Average for group	100.00	98.18	99.89	93.83
Standard Deviation	0.00	2.54	0.35	9.36
Shapiro Wilk Normality Test p-value	$p<0.01$	$p<0.01$	$p<0.01$	$p<0.01$
Data normally distributed?	no	no	no	no

Given that none of the populations contain data that is normally distributed, as determined by a Shapiro-Wilk normality test, analysis of the data with ANOVA or T-Tests is not recommended. A purely descriptive analysis of the averages for the each group suggests that while using non-subjunctive verbs in an S-ONLY context is not exclusively a trait of heritage speakers (since four TJ2 and one SD1 speaker use non-subjunctive in S-ONLY), heritage speakers use non-subjunctive in S-ONLY more than non-heritage speakers. In section 5.5 I provide a logistic regression analysis of subjunctive use in S-ONLY contexts.

5.4.2 Morphosyntactic Categories in S-ONLY contexts

All four morphosyntactic categories of the subjunctive (present, imperfect, present perfect, and pluperfect) are attested in S-ONLY contexts in the border area. In

Table 5.4 I present a breakdown of the morphosyntactic categoriesfor S-ONLY contexts for all four populations. For each group I provide the number of speakers (out of 11) who use that morphosyntactic category and the overall number of instances of that category per group.

Table 5.4 Morphosyntactic Categories in S-ONLY Contexts

		TJ1	TJ2	SD1	SD2
Present Subjunctive	Number of speakers	11	11	10	11
	Total number of instances	144	101	141	90
Present Perfect Subjunctive	Number of speakers	0	0	1	0
	Total number of instances	0	0	1	0
Imperfect Subjunctive	Number of speakers	9	10	9	7
	Total number of instances	69	58	49	13
Pluperfect Subjunctive	Number of speakers	3	2	3	0
	Total number of instances	13	2	3	0

The most common categories in S-ONLY across all populations are present and imperfect subjunctive. The pluperfect subjunctive is used by very few non-heritage speakers and by no heritage speakers in S-ONLY contexts. The present-perfect subjunctive is only attested once by one SD1 speaker in an obligatory context. Considering the low frequency of the present-perfect and pluperfect subjunctive forms, in my analysis of obligatory contexts I exclude the present-perfect subjunctive and combine the pluperfect subjunctive with the imperfect subjunctive since the auxiliary verb *haber* in pluperfect subjunctive is conjugated in imperfect subjunctive. Silva-Corvalán (1994) similarly finds that the present perfect and the pluperfect subjunctive are low frequency even among U.S. immigrant speakers born and raised in Mexico.

Present subjunctive is the most common form in S-ONLY contexts since it is used by 43 out of 44 speakers in the area. Table 5.5 presents the percentages of present subjunctive use in obligatory contexts for all speakers sorted in descending order per group.

Table 5.5 Percentage of Present Subjunctive Verbs in S-ONLY Contexts

Group	TJ1	TJ2	SD1	SD2
Percentage of present subjunctive verbs in S-ONLY contexts = speaker's number of verbs in present subjunctive in S-ONLY / speaker's number of overall tokens in S-ONLY contexts	100.00 100.00 85.00 75.00 64.71 62.96 61.36 54.55 50.00 50.00 44.44	100.00 87.50 75.00 71.43 70.00 64.29 61.90 56.25 50.00 47.62 46.15	100.00 88.89 86.59 80.00 80.00 80.00 66.67 63.64 39.29 30.00 0.00	100.00 100.00 100.00 94.12 92.31 87.50 85.71 80.95 70.00 55.56 40.00
Average for group	68.00	66.38	65.01	82.38
Standard Deviation	18.70	16.08	28.68	18.77
Shapiro Wilk Normality Test p-value	p>0.05	p>0.05	p>0.05	p>0.05
Data normally distributed?	yes	yes	yes	yes

Though a one-way ANOVA ($F(3,40) = 1.453, p = .241$) indicates that there is no statistically significant difference in the percentage of subjunctive verbs used in S-ONLY contexts among the populations, a two-tailed T-Test comparison between heritage and non-heritage speakers ($p<0.040$) suggests that there is a marginal heritage effect among the groups. As a group, the SD2 speakers may show a tendency to use present subjunctive in S-ONLY contexts in a way that may be statistically significant when compared to non-heritage speakers. However, given that the p-value for the T-Test is so

close to the threshold of marginal significance of 0.05, the tendency for SD2 speakers to prefer present subjunctive in S-ONLY contexts can at best and with some reservation be called a trend.

A total of 35 out of 44 speakers use imperfect subjunctive and 8 out of 44 speakers use pluperfect subjunctive. I combine these two morphosyntactic categories for their analysis. It is worth pointing out that all 8 speakers who use pluperfect subjunctive also use imperfect subjunctive. In other words, no speaker uses pluperfect subjunctive only without using imperfect subjunctive. This seemingly systematic pattern of attenuation of the pluperfect subjunctive – in fact, the compound subjunctive tenses – is also attested by Silva-Corvalán (1994) in her Los Angeles data. In Table 5.6 I provide the percentages of imperfect and pluperfect subjunctive use in S-ONLY per speaker as well as an average for each group.

Table 5.6 Percentage of Imperfect & Pluperfect Subjunctive Verbs in S-ONLY Contexts

Group	TJ1	TJ2	SD1	SD2
Percentage of imperfect and pluperfect subjunctive verbs in S-ONLY contexts = speaker's number of verbs in imperfect and pluperfect subjunctive in S-ONLY / speaker's number of overall tokens in S-ONLY contexts	55.56 50.00 50.00 45.45 38.64 37.04 35.29 25.00 15.00 0.00 0.00	53.85 50.00 47.62 43.75 33.33 28.57 28.57 26.67 25.00 12.50 0.00	100.00 70.00 60.71 36.36 33.33 20.00 20.00 12.20 11.11 0.00 0.00	33.33 30.00 30.00 14.29 7.69 5.88 4.76 0.00 0.00 0.00 0.00
Average for group	31.99	31.80	33.06	11.45
Standard Deviation	18.70	15.63	30.24	12.75
Shapiro Wilk Normality Test p-value	$p>0.05$	$p>0.05$	$p>0.05$	$p<0.01$
Data normally distributed?	yes	yes	yes	no

The averages for each of the groups demonstrate that the heritage speakers produce fewer imperfect and pluperfect subjunctives in S-ONLY contexts when compared to their non-heritage counterparts. A one-way ANOVA ($F(3,40) = 2.604$, $p = .065$) indicates no statistically significant difference among the groups. However, an unpaired two-tailed T-Test comparing heritage to non-heritage speakers reveals a strong heritage effect ($p<0.006$). One caveat regarding the T-Test result is that though the data is normally distributed for all non-heritage speakers, it is not so for the heritage speakers since the SD2 sample does not pass the Shapiro Wilk Normality Test. While the T-Test can handle some degree of non-normality, the size of the SD2 sample (11 speakers) may not be large enough in this case. Therefore, I report this result with a caveat and provide a logistic regression analysis of imperfect and pluperfect subjunctive use in S-ONLY contexts in section 5.5.

5.4.3 The Subjunctive in Alternating Contexts

I begin the analysis for subjunctive use in alternating contexts by looking at the proportion of S-ALT contexts that speakers in the border area create, regardless of the mood or morphosyntactic category that they use in the embedded clause. This proportion is calculated by taking into account the number of finite verbs that each speaker produces in the entire interview. Table 5.7 contains the percentages of tokens used in S-ALT contexts for each speaker, sorted in descending order.

Table 5.7 Percentage of S-ALT Contexts

Group	TJ1	TJ2	SD1	SD2
Percentage of S-ALT contexts = speaker's number of tokens in S-ALT contexts / speaker's number of finite verbs	4.22 3.82 2.43 1.77 1.68 1.68 1.56 1.43 0.87 0.29 0.00	5.95 4.17 4.12 3.70 3.50 3.05 3.04 2.86 2.67 2.02 1.52	4.29 2.30 2.15 1.99 1.68 1.13 1.10 0.99 0.91 0.75 0.38	5.08 5.00 2.87 2.19 2.07 1.96 1.69 1.44 1.34 0.93 0.00
Average for group	1.80	3.33	1.61	2.23
Standard Deviation	1.24	1.13	1.03	1.49
Shapiro Wilk Normality Test p-value	p>0.05	p>0.05	p>0.05	p>0.05
Data normally distributed?	yes	yes	yes	yes

Compared to table 5.2, which presented the percentage of S-ONLY contexts, S-ALT contexts are less common for all four populations than S-ONLY contexts. This suggests that all groups generate more S-ONLY contexts than S-ALT contexts. Moreover, the averages for each group indicate that the TJ2 group generates more S-ALT contexts than the other groups. A one-way ANOVA ($F(3,40) = 3.870$, $p = .016$) reveals a marginal statistically significant difference in the number of S-ALT contexts generated among the populations. The results of a post-hoc Tukey HSD test indicate that the statistical significant is due to the TJ2 group generating more S-ALT contexts when compared to TJ1 and SD1 – that is, other non-heritage speakers. An unpaired two-tailed T-Test comparing TJ2 with SD1 and TJ1 reveals a strong statistically significant difference ($p<2.47\text{x}10^{-5}$) for TJ2. Recall from chapter 2 that TJ2 is the group with the highest level

of education in Spanish out of the four populations. It may be that higher education in Spanish is a factor in the use of subjunctive in S-ALT contexts.

By definition, in S-ALT contexts speakers may use a verb in the subjunctive or one not in subjunctive (such as indicative or conditional). In the two tables that follow, I present the percentages of subjunctive verbs (Table 5.8) and the percentages of non-subjunctive verbs (Table 5.9) in S-ALT contexts for each speaker.

Table 5.8 Percentage of Subjunctive Verbs in S-ALT Contexts

Group	TJ1	TJ2	SD1	SD2
Percentage of subjunctive verbs in S-ALT contexts = speaker's number of subjunctive tokens in S-ALT contexts / speaker's number of overall tokens in S-ALT contexts	100.00 92.86 66.67 61.54 57.14 56.00 42.86 40.00 33.33 16.67 0.00	100.00 93.75 91.67 75.00 66.67 65.00 57.14 47.06 40.00 25.00 22.22	100.00 100.00 91.67 70.00 66.67 65.38 57.14 50.00 50.00 20.00 0.00	87.50 80.00 54.55 50.00 50.00 50.00 50.00 33.33 28.57 25.00 0.00
Average for group	51.55	62.14	60.99	46.27
Standard Deviation	28.36	25.60	29.80	23.42
Shapiro Wilk Normality Test p-value	p>0.05	p>0.05	p>0.05	p>0.05
Data normally distributed?	yes	yes	yes	yes

Table 5.9 Percentage of Non-Subjunctive Verbs in S-ALT Contexts

Group	TJ1	TJ2	SD1	SD2
Percentage of Non-Subjunctive verbs in S-ALT contexts = speaker's number of non-subjunctive tokens in S-ALT contexts / speaker's number of overall tokens in S-ALT contexts	83.33 66.67 60.00 57.14 44.00 42.86 38.46 33.33 7.14 0.00 0.00	77.78 75.00 60.00 52.94 42.86 35.00 33.33 25.00 8.33 6.25 0.00	100.00 80.00 50.00 50.00 42.86 34.62 33.33 30.00 8.33 0.00 0.00	75.00 71.43 66.67 50.00 50.00 50.00 50.00 45.45 20.00 12.50 0.00
Average for group	39.36	37.86	39.01	44.64
Standard Deviation	26.33	25.60	29.80	23.10
Shapiro Wilk Normality Test p-value	p>0.05	p>0.05	p>0.05	p>0.05
Data normally distributed?	yes	yes	yes	yes

Neither ANOVAs for differences among groups nor T-Tests for heritage and immigrant comparisons yield any statistically significant difference for the percentage of subjunctive and non-subjunctive verbs used in S-ALT contexts. From a purely descriptive perspective, and only taking into account the average percentages for each group, in S-ALT contexts the heritage speakers are the group with the lowest overall percentage of subjunctive use and with the highest overall percentage of non-subjunctive use. Speakers in the TJ2 group are on the opposite end of the spectrum: in S-ALT contexts they exhibit the overall highest percentage subjunctive use and the lowest overall percentage of non-subjunctive use. None of these observations, however, is statistically significant. However, in spite of no statistically significant difference, TJ2 is the group with the highest educational level in Spanish and SD2 the one with the

lowest. This may indicate that both access to and lack of access to formal schooling in Spanish play a role in mood selection in S-ALT contexts.

5.4.4 Morphosyntactic Categories in S-ALT Contexts

Only the present and imperfect subjunctive feature prominently among the four populations in S-ALT contexts. There are only four instances of present-perfect subjunctive and none of pluperfect subjunctive. Table 5.10 presents a breakdown of the morphosyntactic categories for S-ALT contexts. For each group I provide the number of speakers (out of 11) who use the category and the overall number of instances of that category per group,

Table 5.10 Morphosyntactic Categories in S-ALT Contexts

		TJ1	TJ2	SD1	SD2
Present Subjunctive	Number of speakers	10	11	10	10
	Total number of instances	67	80	58	30
Present Perfect Subjunctive	Number of speakers	0	0	2	1
	Total number of instances	0	0	3	1
Imperfect Subjunctive	Number of speakers	6	6	4	4
	Total number of instances	16	15	7	9
Pluperfect Subjunctive	Number of speakers	0	0	0	0
	Total number of instances	0	0	0	0

Given the lack of instances of pluperfect subjunctive and the extremely low frequency of present-perfect subjunctive, only the present and imperfect subjunctive are considered in the analysis of S-ALT contexts. The (almost complete) absence of the present-perfect and the pluperfect subjunctive in S-ALT contexts in all four groups may indicate that there is a systematic reduction of the compound forms on both sides of the

border even among highly educated monolingual speakers and independent of language contact and language acquisition experience[2].

The present subjunctive is the most common form in S-ALT contexts. Table 5.11 presents the percentages of present subjunctive use in alternating contexts for all speakers, sorted in descending order per group.

Table 5.11 Percentage of Present Subjunctive Verbs in S-ALT Contexts

Group	TJ1	TJ2	SD1	SD2
Percentage of present subjunctive verbs in S-ALT contexts = speaker's number of verbs in present subjunctive in S-ALT / speaker's number of overall tokens in S-ALT contexts	100.00 71.43 63.89 48.00 42.86 40.00 34.62 33.33 28.57 8.33 0.00	81.25 77.78 75.00 57.14 55.00 50.00 41.18 40.00 33.33 25.00 22.22	100.00 75.00 70.00 66.67 59.62 58.33 50.00 42.86 33.33 20.00 0.00	60.00 54.55 50.00 50.00 37.50 28.57 25.00 25.00 25.00 23.81 0.00
Average for group	42.82	50.72	52.35	34.49
Standard Deviation	26.88	19.76	26.34	16.92
Shapiro Wilk Normality Test p-value	p>0.05	p>0.05	p>0.05	p>0.05
Data normally distributed?	yes	yes	yes	yes

A one-way ANOVA ($F(3,40) = 1.284$, $p = .292$) indicates that no statistically significant difference can be found among the populations in the percentage of subjunctive present verbs used in alternating contexts. Two-tailed T-Tests to compare heritage to non-

[2] It may be that these forms are very low-frequency outside of the border area. An investigation on the frequency of these compounds forms in other varieties of Mexican Spanish can help determine whether or not this is the case.

heritage speakers ($p<0.094$) and immigrant to Tijuana speakers ($p<0.558$) show no effect.

Imperfect subjunctive in S-ALT contexts is slightly more common among speakers in Tijuana than among speakers in San Diego. The percentages of imperfect subjunctive use for each speaker in S-ALT contexts are presented in Table 5.12 below.

Table 5.12 Percentage of Imperfect Subjunctive Verbs in S-ALT Contexts

Group	TJ1	TJ2	SD1	SD2
Percentage of imperfect subjunctive verbs in S-ALT contexts = speaker's number of verbs in imperfect and pluperfect subjunctive in S-ALT / speaker's number of overall tokens in S-ALT contexts	28.57 26.92 21.43 8.33 8.00 2.78 0.00 0.00 0.00 0.00 0.00	41.67 33.33 22.22 12.50 10.00 5.88 0.00 0.00 0.00 0.00 0.00	33.33 16.67 14.29 1.92 0.00 0.00 0.00 0.00 0.00 0.00 0.00	50.00 25.00 20.00 9.52 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Average for group	8.73	11.42	6.02	9.50
Standard Deviation	10.88	14.13	10.43	15.44
Shapiro Wilk Normality Test p-value	$p<0.01$	$p<0.01$	$p<0.01$	$p<0.01$
Data normally distributed?	no	no	no	no

None of the four populations pass the Shapiro Wilk Normality Test, which means that the data sample for imperfect subjunctive use in S-ALT contexts does not exhibit normal distribution and therefore an analysis using ANOVA and T-Tests to determine if a statistically significant difference exists is not possible. In section 5.5 of this chapter I provide a logistic regression analysis of the imperfect subjunctive in S-ALT contexts using Rbrul.

5.4.5 The Subjunctive in Variation with the Conditional (COND-EXP)

In the Tijuana-San Diego area 31 out of 44 total speakers use the imperfect and pluperfect subjunctive where a verb in the conditional would be expected. Table 5.13 shows the number of speakers (out of 11) who use the imperfect and pluperfect subjunctive and the overall number of instances of these two forms per group in COND-EXP contexts.

Table 5.13 Subjunctive Morphosyntactic Categories System in COND-EXP

		TJ1	TJ2	SD1	SD2
Imperfect Subjunctive	Number of speakers	5	3	6	9
	Total number of instances	13	5	10	21
Pluperfect Subjunctive	Number of speakers	7	2	4	3
	Total number of instances	38	8	11	10

The group with the most speakers and instances of imperfect subjunctive use is SD2, whereas for pluperfect subjunctive it is TJ1. Unlike the case of S-ONLY contexts – where speakers who use pluperfect subjunctive are a subset within the group of speakers who use imperfect subjunctive – in COND-EXP contexts speakers who use the pluperfect subjunctive do not necessarily use the imperfect subjunctive and vice versa. The use of subjunctive in COND-EXP contexts appears to happening across all populations on both sides of the border.

In Table 5.14 I give the percentages of subjunctive verbs (combined imperfect and pluperfect subjunctive) in COND-EXP contexts for all four populations. For each speaker, the percentages below answer the following question: out of all the verbs in the

subjunctive mood that a speaker produces, what percentage is being used in a COND-EXP context.

Table 5.14 Imperfect and Pluperfect Subjunctive in COND-EXP

Group	TJ1	TJ2	SD1	SD2
Percentage of combined imperfect and pluperfect subjunctive verbs in COND-EXP contexts = speaker's number of imperfect and pluperfect subjunctive verbs in COND-EXP contexts / speaker's overall number of verbs in the subjunctive mood	45.45 28.57 26.67 22.22 20.83 15.69 15.00 0.00 0.00 0.00 0.00	23.81 10.00 9.68 9.09 4.35 0.00 0.00 0.00 0.00 0.00 0.00	27.27 23.53 21.21 15.38 11.11 7.14 3.45 1.74 0.00 0.00 0.00	50.00 41.18 36.36 28.57 25.00 25.00 20.00 17.39 12.00 6.25 0.00
Average for group	15.86	5.18	10.08	23.80
Standard Deviation	14.22	7.18	9.80	14.27
Shapiro Wilk Normality Test p-value	$p<0.01$	$p<0.01$	$p<0.01$	$p<0.01$
Data normally distributed?	no	no	no	no

Given the low frequency for both imperfect and pluperfect subjunctive, none of the percentages of subjunctive use in COND-EXP for any of the non-heritage populations pass the Shapiro-Wilk Normality Test. As such, neither the ANOVA nor the T-Test are recommended for data that is not normally distributed but the logistic regression analysis for data that is not normally distributed is presented in section 5.5.

The group averages from Table 5.14 indicate that heritage speakers use imperfect and pluperfect subjunctive verbs outside of subjunctive contexts (either S-ONLY or S-ALT) more than any other group. This may suggest the following about

heritage speakers' use of imperfect and pluperfect subjunctive if a statistically significant difference can be determined:

1. Heritage speakers reduce their use of imperfect and pluperfect subjunctive verbs in comparison to non-heritage speakers in S-ONLY contexts (recall the results from Table 5.3.
2. Heritage speakers use more imperfect and pluperfect subjunctive verbs in non-subjunctive contexts (i.e., COND-EXP) than non-heritage speakers.

In the following section 5.5 I provide a logistic regression analysis of subjunctive use with Rbrul, which can handle data that is not assumed to have normal distribution.

5.5 Logistic Regression Analysis of the Subjunctive

Rbrul is an R-based platform that can perform regression analysis on data that is "unbalanced across the factors of interest" and can "estimate the effects of multiple predictors" (Johnson 2009). Recall from the previous section that the unbalanced data for the imperfect and pluperfect subjunctive in S-ALT and COND-EXP contexts does not pass the Shapiro Wilk test for normality. As such, a statistically significant difference cannot be determined by parametric tests such as ANOVA and the T-Test, but Rbrul makes it possible to analyze this unbalanced data. Another advantage that Rbrul presents is that it controls for Type I errors – that is, where significance may be overestimated by wrongly identifying a chance effect as a real effect. The flipside of that is that it may

instead, in some cases, generate a Type II error where it fails to identify an effect that may be significant (Johnson 2009).

In this study of subjunctive across four populations in the border area, there are two types of predictors for subjunctive mood use: internal and external. Table 5.15 outlines the internal and external predictors.

Table 5.15 Internal and External Predictors for Subjunctive Use

Type of Predictor	Predictor	Possible values
Internal	Mood	Subjunctive, Indicative
	Tense	Present, Imperfect, etc.
	Context	S-ONLY, S-ALT, COND-EXP
External	Speaker	TJ-S1, SD-B2, etc
	Group	TJ, SD
	Gender	M, F
	Heritage	Heritage, Non-heritage
	Generation	1, 2
	Occupation	Business owner, student, etc
	Educational level	High school finished, etc
	Years in the border	18-45
	Socioeconomic status	Working class, professional, etc

The previous section (section 5.4) contains a complete overview of the internal predictors of subjunctive use. The chapter on populations (Chapter 2) provides a thorough discussion of the external sociodemographic predictors as well as a complete list of all the actual values for each predictor for the four populations in the border. In the subsections that follow I present the results and analysis of six logistic regression runs that correspond to mood selection and the morphosyntactic categories of the subjunctive. Since the goal of this project is to determine differences in mood selection

on both sides of the border, the predictor GROUP is included in every logistic regression run. If the predictor GROUP emerges as significant and the group SD2 is revealed to be significantly different from the other groups, I perform the regression run with the predictor heritage to determine if HERITAGE is a stronger predictor than GROUP. In such case, I determine whether there is a heritage effect.

5.5.1 Predictors of General Subjunctive Use in the Border Area

The first Rbrul run analyzes MOOD as the dependent variable with the application value of SUBJUNCTIVE. As shown in Table 5.16, 1176 out of a total of 1292 were analyzed. These 1176 tokens correspond to all instances of verbs in embedded clauses for S-ONLY and S-ALT contexts. The remaining 116 verbs belong to COND-EXP contexts and as such their ratio of subjunctive use is 1.000 since by definition the COND-EXP context refers to verbs used in imperfect and pluperfect subjunctive where a conditional is expected. I exclude COND-EXP contexts from this run and analyze them in a separate run.

Table 5.16 One-level Analysis of Subjunctive Use in the Border Area

	Mood: Subjunctive			
Number of raw tokens	1176			
	Log odds	SUBJ/ SUBJ + IND	Number of tokens	Factor weight
Context	**p.<2.88e-71**			
S-ONLY	1.883	0.981	697	0.868
S-ALT	-1.883	0.597	479	0.132
Heritage	**p.<0.0369**			
Non-heritage	0.378	0.981	983	0.593
Heritage	-0.378	0.741	193	0.407

The strongest significant predictor for subjunctive use is CONTEXT (p.<2.88e-71). Positive log odds for the S-ONLY context indicate a positive correlation between this factor and the dependent variable (MOOD: SUBJUNCTIVE). The high factor weight of 0.868 for S-ONLY indicates two things: (1) when a verb in the subjunctive occurs, it is more likely to occur in an S-ONLY context than in an S-ALT context, which has a low factor weight of 0.132 and negative log odds; (2) the factor weight is sufficiently removed from the neutral mark of 0.5 for factor weights, which indicates a greater size for the effect.

The second strongest predictor is HERITAGE group (p.<0369), with non-heritage speakers having a positive correlation (log odds) with the subjunctive mood. A verb in the subjunctive is more likely to be used by non-heritage speakers than by heritage speakers. However, the effect size is not as great as it is for CONTEXT given that the factor weights for both heritage and non-heritage speakers are close to the neutral mark of 0.5. The rationale for including a predictor for heritage group that combines the SD1, TJ1 and TJ2 populations is that the Rbrul run with GROUP as a predictor (shown in Table

5.17) suggests that the SD2 group behaves differently from the other three groups. The SD2 group is the only one with negative log odds and a factor weight that deviates from the 0.5 mark more than any of the other three groups. Even though GROUP is non-significant (p<0.182) as a predictor, combining SD1, TJ1 and TJ2 into a non-heritage group for comparison with the heritage group (SD2) reveals a significant HERITAGE effect rather than a GROUP effect.

Table 5.17 GROUP as a Non-significant Predictor for Subjunctive Mood

Group	**[p.<0.182]**			
	Log odds	SUBJ/ SUBJ + IND	Number of tokens	Factor weight
SD1	0.351	0.876	299	0.587
TJ2	0.187	0.813	315	0.547
TJ1	0.038	0.837	369	0.509
SD2	-0.576	0.741	193	0.360

One internal (CONTEXT) and one external predictor (HERITAGE) are the only ones that are statistically significant for subjunctive use in the border area. The results of this run indicate that when a verb occurs in the subjunctive in the border area, it is significantly more likely to have been produced in an S-ONLY context than in an S-ALT context. It is also more likely to have been produced by a non-heritage speaker than by a heritage speaker. I avoid including TENSE as a predictor in this first run for overall subjunctive use for two reasons: (1) I perform separate individual runs for present and imperfect subjunctive; (2) the subjunctive mood only has two morphosyntactic categories whereas the indicative mood has many. With TENSE as a stand-alone

predictor, Rbrul would factor in values such as 'preterite' and 'future' that only pertain to the indicative but not the subjunctive and therefore skew the results.

External predictors other than GROUP that are not statistically significant include GENERATION ($p<0.446$), GENDER ($p<0.359$), SOCIOECONOMIC STATUS ($p<0.282$), OCCUPATION ($p<0.289$), and EDUCATION ($p<0.0824$). None of these sociodemographic factors are significant predictors of subjunctive use in the border. Excluded from the run were the external sociodemographic predictors AGE and YEARS IN THE BORDER since their range of variation is already encompassed by the predictor GENERATION.

5.5.2 Mood Selection: Predictors of Use of Subjunctive in S-ONLY Contexts

The data presented in Table 5.3 for subjunctive use in S-ONLY contexts failed to pass the Shapiro Wilk Normality Test. For this logistic regression analysis I exclude the 116 verbs that belong to COND-EXP contexts and only consider for analysis the 1176 for S-ONLY and S-ALT contexts. I also exclude TENSE as a predictor for both S-ONLY and S-ALT runs since I perform individual runs for the two morphosyntactic categories of the subjunctive. Furthermore, by including individual speakers variation (predictor SPEAKER) as a random effect strengthens the confidence in the significance of other factors given the non-normal distribution of the data. The results are in Table 5.18 below.

Table 5.18 One-level Analysis of Subjunctive Use in S-ONLY Contexts

	Context: S-ONLY			
Number of raw tokens	1176			
	Log odds	S-ONLY/ S-ONLY + S-ALT	Number of tokens	Factor weight
Mood	**p.<9.46e-71**			
Subjunctive	1.873	0.705	970	0.867
Indicative	-1.873	0.063	206	0.133
Group	**[p.<0.108]**			
SD2	0.158	0.575	193	0.540
SD1	0.150	0.652	299	0.537
TJ1	0.144	0.612	369	0.536
TJ2	-0.453	0.524	315	0.389
Speaker	**random, not tested**			
	0.492	0.593	1176	…

The only significant predictor for S-ONLY contexts in the border area is MOOD. In an S-ONLY context the subjunctive mood is significantly more likely to occur given its high factor weight of 0.867. The indicative mood is a weak predictor and highly unlikely to occur given its negative correlation (log odds) to the dependent variable and its low factor weight of 0.133. GROUP does not emerge as s significant predictor (p.<0.108), which suggests that in terms of S-ONLY contexts all four populations do not behave significantly different from one another. Grouping SD1, TJ1 and TJ2 in a predictor HERITAGE would not likely result in an effect since the factor weights for SD1, TJ1 and SD2 is so close to the neutral mark of 0.5, which suggests that the heritage speakers do not perform all that differently from SD1 and TJ1. In Table 5.19 I provide the predictor HERITAGE as part of the Rbrul run to show that no statistically significant difference emerges from this comparison.

Table 5.19 HERITAGE as a Non-significant Predictor for Subjunctive Mood

Heritage	**[p.<0.427]**			
	Log odds	SUBJ/ SUBJ + IND	Number of tokens	Factor weight
Heritage	0.114	0.575	193	0.529
Non-Heritage	-0.114	0.596	983	0.471

In sum, the populations are not significantly different from one another in S-ONLY contexts.

5.5.3 Mood Selection: Predictors of Use of Subjunctive in S-ALT Contexts

As with the run for S-ONLY contexts, this run excludes the 116 subjunctive verbs that occur in COND-EXP contexts. I also do not test for TENSE as a predictor since I perform separate runs for the two morphosyntactic categories of the subjunctive. Speakers are included as a random effect with the predictor SPEAKER to take into account individual speaker variation in data that is not normally distributed. Table 5.20 outlines the results of the one-level analysis of subjunctive use in S-ALT contexts.

Table 5.20 One-level Analysis of Subjunctive Use in S-ALT Contexts

	Context: S-ALT			
Number of raw tokens	1176			
	Log odds	S-ALT/ S-ONLY + S-ALT	Number of tokens	Factor weight
Mood	**p.< 9.46e-71**			
Indicative	1.873	0.937	206	0.867
Subjunctive	-1.873	0.295	970	0.133
Group	**[p.<0.108]**			
TJ2	0.453	0.476	315	0.611
TJ1	-0.144	0.388	369	0.464
SD1	-0.150	0.348	299	0.463
SD2	-0.158	0.425	193	0.461
Speaker	**random, not tested**			
	0.492	0.407	1176	…

Indicative mood is the strongest predictor of S-ALT contexts with a factor weight of 0.867 and a positive correlation (log odds) with the dependent variable (context: S-ALT). The effect size is considerable given that the factor weight is far above the neutral mark of 0.5. In S-ALT contexts the subjunctive mood is a significantly weak predictor with a factor weight of 0.133. GROUP does not emerge as a significant predictor for S-ALT contexts. The factor weights for the four populations reveal they perform very closely to the neutral mark of 0.5. This suggests that in S-ALT contexts the four groups do not perform statistically differently from one another. The predictor HERITAGE is not considered in this run since the only group that performs differently from the rest is TJ2 (positive log odds, higher factor weight).

Recall from Tables 5.7 and 5.8 that the TJ2 group exhibits a higher percentage use of subjunctive contexts and subjunctive verbs in S-ALT contexts. In this one-way analysis, TJ2 emerges as the only group with a positive correlation to the dependent

variable S-ALT. The lack of a statistically significant difference notwithstanding, the TJ2 speakers exhibit different behavior from the other groups in S-ALT contexts which may be related to their higher educational level in Spanish.

5.5.4 Morphosyntactic Categories: Predictors of Use of Present Subjunctive

For the logistic regression analyses of the t7wo forms of the subjunctive (present and imperfect subjunctive) I combine MOOD and TENSE to create the dependent variable VERB with the application value of SUBJPRES. The 1176 verbs from the previous runs are part of the analysis. For the predictor CONTEXT, the value COND-EXP is excluded from this analysis since by definition all verbs in COND-EXP are imperfect and pluperfect subjunctive. Speakers are also included as a random effect in this run. Table 5.21 presents the results of this one-level analysis.

Table 5.21 One-level analysis of the context SUBJPRES

	Verb: SUBJPRES			
Number of raw tokens	1176			
	Log odds	SUBJPRES/ ALL TENSES	Number of tokens	Factor weight
Context	**p.< 6.39e-12**			
S-ONLY	0.446	0.686	697	0.610
S-ALT	-0.446	0.499	479	0.390
Group	**[p.< 0.636]**			
SD1	0.148	0.682	299	0.537
SD2	0.132	0.627	193	0.533
TJ2	-0.103	0.575	315	0.474
TJ1	-0.177	0.572	369	0.456
Speaker	**random, not tested**			
	0.527	0.61	1176	…

CONTEXT is the only significant predictor for use of present subjunctive. When a verb occurs in the present subjunctive is more likely to be in in an S-ONLY context given the positive correlation (log odds) with the dependent variable (SUBJPRES). The S-ALT context is also a significant factor in that a present subjunctive verb is less likely to be in an S-ALT context. GROUP does not emerge as a significant predictor. Grouping the populations into heritage and non-heritage groups to test for a HERITAGE effect proves to also not be significant as shown in Table 5.22.

Table 5.22 HERITAGE as a Non-significant Predictor for Present Subjunctive

Heritage	**[p.<0.473]**			
	Log odds	SUBJPRES/ ALL TENSES	Number of tokens	Factor weight
Heritage	0.096	0.627	193	0.524
Non-Heritage	-0.096	0.606	983	0.476

The lack of a group and heritage effect demonstrate that the four groups do not perform significantly different from each other in their use of the present subjunctive.

5.5.5 Morphosyntactic Categories: Predictors of Imperfect and Pluperfect Subjunctive

As I did for the logistic regression run on present subjunctive, for this run on imperfect subjunctive use (which includes the pluperfect subjunctive), I combine MOOD and TENSE to create the dependent variable VERB with the application value of SUBJIMPERF. The same 1176 verbs from the previous runs are part of the analysis. For the predictor CONTEXT, the value COND-EXP is excluded from this analysis since by

definition all verbs in COND-EXP are imperfect and pluperfect subjunctive. The one-level analysis, with individual speakers included as a random effect, is presented in Table 5.23.

Table 5.23 One-level Analysis of the Imperfect (and Pluperfect) Subjunctive

	Verb: SUBJIMPERF			
Number of raw tokens	1176			
	Log odds	SUBJIMPERF/ ALL TENSES	Number of tokens	Factor weight
Context	**p.<1.56e-15**			
S-ONLY	0.68	0.296	697	0.664
S-ALT	-0.68	0.098	479	0.336
Heritage	**p.<0.01**			
Non-heritage	0.484	0.235	983	0.619
Heritage	-0.484	0.114	193	0.381
Speaker	**random, not tested**			
	0.725	0.215	1176	…

Within the two significant predictors CONTEXT and HERITAGE, the factors with a positive correlation (log odds) to imperfect subjunctive use are S-ONLY and NON-HERITAGE. The factor weights (0.664 and 0.619, respectively) above the neutral mark of 0.5 indicate that when the imperfect subjunctive is used, it is more likely to be in an S-ONLY context and by a NON-HERITAGE speaker. The factors with negative correlation (log odds) – S-ALT contexts and HERITAGE speakers – have factor weights (0.336 and 0.381, respectively) below the neutral mark of 0.5. This suggests that when an imperfect and/or pluperfect subjunctive verb occurs, it is less likely to be in an S-ALT context by a HERITAGE speaker.

5.5.6 Morphosyntactic Categories: Predictors of Imperfect and Pluperfect Subjunctive Use in COND-EXP Contexts

From the previous logistic regression runs I have excluded the 116 tokens in COND-EXP contexts since they only occur with the imperfect and pluperfect subjunctive. I now include them as part of the analysis but forego TENSE as a predictor since all the tokens in COND-EXP contexts occur in the imperfect or pluperfect subjunctive. I combine the pluperfect and imperfect subjunctive for analysis given their low frequency and the fact that the auxiliary verb *haber* for pluperfect subjunctive is in the imperfect subjunctive. The results of this one-level analysis with speakers as a random effect are presented in Table 5.24.

Table 5.24 One-level Analysis of the Context COND-EXP

	Context: COND-EXP			
Number of raw tokens	1292			
	Log odds	COND-EXP/ ALL CONTEXTS	Number of tokens	Factor weight
Group	**p.<0.0105**			
SD2	0.658	0.138	224	0.659
TJ1	0.331	0.121	420	0.582
SD1	-0.090	0.066	320	0.477
TJ2	-0.899	0.040	328	0.289
Speaker	**random, not tested**			
	0.705	0.09	1292	…

The strongest predictor is GROUP but the statistically significant difference is low (p.<0.0105). The heritage speakers (SD2) have a marked preference for COND-EXP contexts while the TJ2 speakers are less likely to use this context. This is evidenced by

the deviation of the factor weights from the neutral mark of 0.5 for the SD2 and TJ2 groups. Given that the TJ2 group behaves differently from the rest of the non-heritage speakers (SD1 and TJ1), combining the three non-heritage groups into a single non-heritage group for a predictor heritage results in a weaker effect (p.<0.0319), as shown in Table 5.25.

Table 5.25 HERITAGE as a Predictor for COND-EXP Contexts

Heritage	**p.<0.0319**			
	Log odds	COND EXP/ ALL CONTEXTS	Number of tokens	Factor weight
HERITAGE	0.425	0.138	224	0.605
HON-HERITAGE	-0.425	0.080	1068	0.395

These results indicate that when a verb occurs in a COND-EXP context, it is more likely to be used by a SD2 speaker and quite unlikely to be used by a TJ2 speaker. Access to formal schooling in Spanish may be playing a role since SD2 and TJ2 are located at opposite ends of the spectrum of education in Spanish.

5.5.7 Summary of Findings: Mood Selection and the Morphosyntactic Categories

In Table 5.26 below I present the significant results from the regression runs. For each dependent variable, the significant predictors are accompanied by its p-value. The values with the strongest effect size for each predictor are accompanied by their factor weights.

Table 5.26 Significant Results of Regression Runs

	Dependent Variable	Predictors (p-value)	Values (fw > 0.5)	Values (fw < 0.5)
MOOD	SUBJUNCTIVE	CONTEXT (p< 2.88e-71)	S ONLY (fw 0.868)	S-ALT (fw 0.132)
		HERITAGE (p<0.0369)	NON-HERITAGE (fw 0.593)	HERITAGE (fw 0.407)
MOOD SELECTION	S-ONLY	MOOD (p<9.46e-71)	SUBJUNCTIVE (fw 0.867)	INDICATIVE (fw 0.133)
	S-ALT	MOOD (p<9.46e-71)	INDICATIVE (fw 0.867)	SUBJUNCTIVE (fw 0.133)
TENSE	PRESENT SUBJUNCTIVE	CONTEXT (p<6.39e-12)	S ONLY (fw 0.610)	S-ALT (fw 0.390)
	IMPERFECT & PLUPERFECT SUBJUNCTIVE	CONTEXT (p<1.56e-15)	S ONLY (fw 0.664)	S-ALT (fw 0.336)
		HERITAGE (p<0.01)	NON-HERITAGE (fw 0.619)	HERITAGE (fw 0.381)
OTHER	COND-EXP	GROUP (p<0.0105)	SD2 (fw 0.659)	TJ2 (fw 0.289)

Speaking of the subjunctive in general terms, the subjunctive mood is low frequency in the Tijuana – San Diego border area with only 5.3% of finite verbs produced in the subjunctive. A verb in the subjunctive is more likely to occur in an S-ONLY context than in an S-ALT context, and it is slightly more likely to be produced by a non-heritage speaker than by a heritage speaker. The statistically significant difference of this heritage effect is small. It is also worth pointing out that all four populations exhibit a high degree of intra-group variation in the amount of subjunctive verbs that speakers produce. No sociodemographic factors emerge as significant for subjunctive use in the border area.

In terms of mood selection, in S-ONLY contexts the subjunctive mood is significantly more likely to occur. In S-ALT context, it is the indicative mood that is the

strongest predictor. None of the four populations behave any differently from one other with respect to mood selection though TJ2 speakers appear to create more S-ALT contexts than any other group even though this is not statistically significant.

For the morphosyntactic forms of the subjunctive mood, S-ONLY contexts are strong predictors of both the present and the imperfect/pluperfect subjunctive. There is no statistically-significant difference among the four groups in their use of subjunctive present. However, if a verb occurs in imperfect or pluperfect subjunctive it is somewhat more likely to be produced by a non-heritage speaker than by a heritage speaker. The low frequency of the pluperfect subjunctive in relation to the imperfect subjunctive in subjunctive contexts (either S-ONLY or S-ALT) may be part of a pattern of loss and simplification of the tense-mood-aspect-system wherein the compound forms are more vulnerable to loss than the simple forms, as Silva-Corvalán (1994) finds in Los Angeles Spanish.

The use of the imperfect and pluperfect subjunctive in place of the conditional in non-subjunctive contexts is more likely to occur among SD2 heritage speakers. Speakers in the TJ2 group are the least likely to use the imperfect and pluperfect subjunctive in COND-EXP contexts. This use of imperfect and pluperfect subjunctive in place of the conditional is ongoing on both sides of the border across generations but significantly more so among heritage speakers in San Diego. Moreover, this use of the imperfect and pluperfect subjunctive forms does not conform to the systematic attenuation pattern that imperfect and pluperfect subjunctive exhibit in subjunctive contexts. In COND-EXP contexts, some speakers use the pluperfect subjunctive and not

the imperfect subjunctive and vice versa. I dedicate the following section to the use of the subjunctive in COND-EXP contexts.

5.6 A Closer Look at the Imperfect and Pluperfect Subjunctive in COND-EXP Contexts

Sections 5.5.1-5.5.6 of my analysis of mood selection focused on the 1176 tokens used in S-ONLY and S-ALT contexts. The 116 tokens where the subjunctive is used in COND-EXP contexts correspond to 9% of the total corpus of 1292 tokens. Given that the use of subjunctive in COND-EXP contexts is attested in all four population, that it is statistically significant for heritage speakers (who emerge as the strongest factor in the GROUP predictor), and that it accounts for almost 10% of subjunctive use in the border, I dedicate this section to this use of the subjunctive.

I run two separate Rbrul runs with the combined imperfect and pluperfect subjunctive as the dependent variable SUBJIMPERF. The first (Table 5.27) includes only the 1176 S-ONLY and S-ALT contexts for mood selection and only HERITAGE as a predictor,[3] which is a significant one even if the effect is low.

[3] Recall from table 5.23 that with the dependent variable SUBJIMPERF the predictor CONTEXT emerged as a strong predictor. This new Rbrul run excludes CONTEXT as a predictor to highlight the HERITAGE effect.

Table 5.27 One-level Analysis of the Imperfect (and Pluperfect) Subjunctive, excluding COND-EXP

	Verb: SUBJIMPERF			
Number of raw tokens	1176			
	Log odds	SUBJIMPERF/ ALL TENSES	Number of tokens	Factor weight
Heritage	**p.<0.0105**			
Non-heritage	0.487	0.235	983	0.619
Heritage	-0.487	0.114	193	0.381
Speaker	**random, not tested**			
	0.725	0.215	1176	…

Adding to the run the additional 116 tokens that correspond to the COND-EXP contexts (Table 5.28) renders the HERITAGE effect non-existent.

Table 5.28 One-level Analysis of the Imperfect (and Pluperfect) Subjunctive, including COND-EXP

	Verb: SUBJIMPERF			
Number of raw tokens	1292			
	Log odds	SUBJIMPERF/ ALL TENSES	Number of tokens	Factor weight
Heritage	**p.<0.293**			
Non-heritage	0.153	0.296	1068	0.538
Heritage	-0.153	0.237	224	0.462
Speaker	**random, not tested**			
	0.621	0.286	1292	…

This may suggest that while heritage speakers demonstrate an attenuated use of imperfect subjunctive in S-ALT contexts (see Table 5.23), the high frequency of imperfect and pluperfect subjunctive verbs that they demonstrate in COND-EXP contexts compensates for this attenuation. In other words, heritage speakers do not reduce their

overall use of imperfect and pluperfect subjunctive. They reduce it only in S-ALT contexts and increase it in non-subjunctive COND-EXP contexts to such an extent that there is no statistically significant difference among the groups. Further investigation into the types of verbs and clauses used by heritage speakers in COND-EXP contexts can help determine if this may be a case of what Lynch (1999) calls unsystematic use of mood – meaning that the subjunctive and the conditional are used interchangeably with no regard for use or meaning.

5.7 Mood Selection and the Morphosyntactic Categories of Subjunctive in the Border Area

I now revisit the four research questions that guide this investigation of subjunctive in the border area and provide a summary of the main findings.

1. What is the frequency of the subjunctive in the border area and how does it compare to other varieties of Spanish in the U.S.? Do any external sociodemographic factors play a significant role in any difference n subjunctive use?

Collentine (2010) posits that the subjunctive in all of its forms corresponds to approximately 7.2% of all verb forms in Spanish. Bookhamer (2013) finds that in New York immigrant speakers' subjunctive verbs comprise 6.8% of overall verb forms. This number is 5.2% for heritage speakers. Geeslin & Gudmestad (2010) observe that percentage of subjunctive verbs in adult immigrant speakers in the U.S. is about 6.7%.

My findings indicate that in San Diego the percentage of subjunctive verbs for first-generation speakers in 4.75%, for second generation speakers it is 4.82%. In Tijuana, the first generation uses 4.86% of overall verbs in the subjunctive while the second generation uses 5.98%. Generally speaking, use of subjunctive in the Tijuana-San Diego border area is lower when compared to Collentine's (2010) and Bookhamer's (2013) findings. The only sociodemographic predictor for subjunctive use in the Tijuana-San Diego border area is heritage and its statistically significant difference is marginal: heritage speakers as a group show a slight reduction in frequency of subjunctive use.

2. Are there any statistically significant differences in mood selection in S-ONLY and S-ALT contexts among the four populations?

There are no statistically significant differences in mood selection in S-ONLY and S-ALT contexts among the four populations. However, there are internal predictors that demonstrate a strong statistically significant difference. In S-ONLY contexts the stronger predictor is the subjunctive mood while in S-ALT context the strongest predictor is the indicative mood.

3. Are the morphosyntactic categories of the subjunctive stable for all four population or do any of the forms undergo simplification or attenuation?

The present subjunctive is stable across all four populations. Heritage speakers exhibit an attenuation of the imperfect subjunctive in S-ALT contexts. The statistically significant difference for this heritage effect is low.

4. Given that the subjunctive mood is extended to contexts where the conditional is expected, is this change systematic and does it apply to all four populations in the border differently?

All four populations in the Tijuana-San Diego border area extend the subjunctive mood, specifically the imperfect and pluperfect subjunctive, in COND-EXP contexts. Heritage speakers, however, do so to an extent that is statistically significant. This overextension of the imperfect and pluperfect subjunctive in COND-EXP contexts on the part of the heritage speakers makes them appear almost indistinguishable from non-heritage speakers in terms of overall imperfect/pluperfect subjunctive use.

The lack of a stronger statistically-significant effect for overall subjunctive use in the heritage population calls into question the notion of simplification of the subjunctive system in the border area. Rather than simplification, what may be happening among heritage speakers in San Diego is attenuation only in S-ALT contexts. It is worth pointing out that extension of the indicative and attenuation of the subjunctive in S-ALT contexts does not result in ungrammaticality, as Silva-Corvalán (1994) points out. Language contact with English can also be discarded as a significant factor given that in terms of interaction and exposure the heritage population is in constant contact with English on a daily basis, and yet it appears to have no great effect on the subjunctive system. Similarly, attrition as a possible cause for the attenuation of subjunctive in S-ALT contexts among heritage speakers can only be ascertained with longitudinal data.

In the case of San Diego heritage speakers, it is more likely that a heritage mode of acquisition of the subjunctive mood results in high variability and attenuation of the

imperfect and pluperfect subjunctive in S-ALT contexts. In describing the acquisition of subjunctive among monolingual speakers, Blake (1985) notes that some obligatory contexts are acquired first, specifically those expressing volition (also Gudmestad 2008), indirect commands, and anticipation. Between the ages of 5 and 8, children go through a transitional period of experimentation in both S-ALT and S-ONLY contexts. Children do not reach significant improvement in S-ALT contexts, particularly in adjectival clauses, until age 9. In a heritage mode of acquisition, the onset of English education and bilingualism coincides precisely with the age at which children are acquiring S-ALT contexts. Therefore, it is possible that heritage speakers continue to fine-tune subjunctive use in S-ALT contexts during adolescence and into adulthood. Since the attenuation of subjunctive and the overextension of indicative do not result in ungrammatical forms in S-ALT contexts, heritage speakers exhibit great variability in these contexts depending on the types of matrices and clauses that they have been exposed to. Blake (1985) also points out that the subjunctive is acquired gradually in childhood and that "exposure to individual lexical items" helps in the acquisition of mood selection.

It is important to emphasize that a heritage mode of acquisition is not one that is incomplete. Blake (1985) notes that even among monolingual speakers the development of subjunctive fails to conform to a single pattern and that sociolinguistic pressures may lead speakers to match certain prescriptive tenets. The lack of formal schooling in Spanish removes these sociolinguistic pressures for many heritage speakers and the development of mood selection in S-ALT contexts is very much dependent on changes

in each heritage speaker's linguistic environment. Collentine (1995) mentions that given the low frequency of subjunctive even in non-contact varieties of Spanish, what heritage speakers in a language contact situation encounter may be even lower. Considering that TJ2 and SD2 speakers differ in their production of subjunctive in S-ALT contexts just as they are on opposite ends on the spectrum of access to formal schooling in Spanish, it may be that acquiring the subjunctive in an academic environment exposes TJ2 speakers to a greater number of contexts for subjunctive use in S-ALT contexts. San Diego second-generation speakers, on the other hand, not only lack access to formal schooling in Spanish but they are also English-dominant in their interactions outside the home.

The overextension of pluperfect and imperfect subjunctive in COND-EXP contexts that San Diego heritage speakers exhibit is not a result of a heritage mode of acquisition. Because this overextension is firmly present among non-heritage populations, heritage speakers acquire it in much the same way TJ2 speakers would. Silva-Corvalán (1994) adds that the subordinate status of Spanish as a community or family language in the U.S. accelerates in heritage speakers an extension process that is caused by a language-internal loss and simplification of the conditional form that is already ongoing in the Spanish-speaking world. Consider the utterance in (17), given by an adult immigrant speaker in San Diego who uses the conditional (*encantaría*) in the first apodasis and the imperfect indicative (*lo regresaba*) in the second apodasis even though the protasis is the same and the two hypothetical conditional sentences are uttered back to back by the same speaker.

(17) *Si pudiera regresar el tiempo me encantaría. Si pudiera regresar el tiempo lo regresaba* [SD-B1]

'If I could turn back time, I would love to. If I could turn back time, I'd turn it back'

What may be occurring in utterances like (17) is what Lynch (1999) calls an unsystematic use of mood wherein the subjunctive and the indicative have lost their distinction.

Lack of education in Spanish may be a factor in heritage speakers' mood selection not only in S-ALT contexts but also in COND-EXP contexts. In Table 5.24 the one-level analysis of COND-EXP contexts revealed a group effect with SD2 as the strongest factor and TJ2 as the weakest. This may indicate that in addition to creating an environment that favors the creation of S-ALT contexts and the use of subjunctive in these contexts, formal schooling in Spanish may also decelerate the process of overextension of the subjunctive into non-subjunctive contexts, namely COND-EXP contexts, since TJ2 speakers show a negative correlation with these contexts that is different in a way that is statistically significant.

Lack of access to formal schooling in Spanish is not the only external factor that results in some differences in subjunctive use among heritage speakers in San Diego. The SD2 speakers are also English-dominant in their interactions outside of the home in addition to having no formal schooling in Spanish. The results of this study show that the onset of schooling in English coincides with a transitional period in the acquisition of the Spanish subjunctive that is characterized by fluctuation and experimentation.

Language dominance in English and lack of access to education in Spanish results in a different use of subjunctive among heritage speakers wherein they reduce the overall frequency of subjunctive in subjunctive contexts while at the same time increase the frequency of imperfect subjunctive in COND-EXP contexts. It is worth pointing out that the use of imperfect subjunctive in COND-EXP contexts is a feature of the Spanish spoken on both sides of the border. However, this feature becomes more prevalent in a contact situation with English that is characterized by heritage speakers being dominant in English and having no formal education in Spanish.

Chapter 6 Fillers

6.1 Fillers in Spanish

In the study of spoken language production, fillers consist of pauses, discourse markers, and repetitions. Pauses can be either filled or silent, and a filled pause can consist of a word or a sound (Watanabe et al 2008). In Spanish, a filled pause is a word like *este* or non-lexical formations such as *ah* and *eh*. Both of these filled pauses, *este* and *eh*, are attested in the same sentence in (1)[1].

(1) *Bueno, la primera es mi familia, eh, Dios, este, y creo que es todo.*

'well, first it's my family, *eh*, God, *este*, and I think that's all'.

[TJ-B1]

Discourse makers in Spanish such as *como que*, *bueno*, and *pues* provide structure to utterances without interfering with their meaning or grammaticality. In (1) the discourse marker *bueno* indicates that the speaker is about to begin a discourse segment. Repetitions refer to any word or filler that is repeated, as well as to false starts. This study focuses only on the use and variety of filled pauses and discourse markers. For a study that analyzes repetitions and silent pauses in addition to filled pauses and discourse markers, see Valdés & Geoffrion-Vinci (1998).

[1] I do not provide an English gloss for the fillers since the English equivalent may not convey the same functions of the Spanish original.

6.2 The Importance of Fillers

Fillers have been an important part of the study of bilingualism and languages in contact. They have been suggested as a useful tool in the measurement of oral proficiency in heritage speakers and language learners (Segalowitz 2004). They have also been partially correlated with verbal fluency in Spanish-English bilinguals (Tavison 2014). Moreover, Spanish-English bilinguals who use English discourse markers more than Spanish ones may be perceived as being less proficient in Spanish (Vickers & Goble 2011). Because some discourse markers can be multifunctional, populations may use them differently. The multifunctionality of discourse markers can vary from one Spanish-English bilingual system to another depending on the language contact situation wherein similar discourse markers in Spanish and English develop differentiated meanings when they coexist (Torres & Potowski 2008). Furthermore, discourse markers, in particular, can also point to a speaker's sensitivity to sociocultural knowledge (Gumperz 1992). In terms of language identity, in informal speech in Spanish – as it is the case for the corpus in this current study – the use of English discourse markers may represent a way for speakers to assert their dual language identities (Zentella 1997).

6.3 Background

Previous research on fillers in Spanish in the U.S. has focused on describing the systems of fillers among Spanish-English bilinguals, and the use of English discourse markers in U.S. Spanish, particularly *so*.

In a study that compares monolingual speakers of Spanish at a Mexican university with bilingual Chicano speakers at a U.S. university, Valdés & Geoffrion-Vinci (1998) find that the most common fillers among both monolingual and Chicano speakers are *bueno*, *pues*, and *este*. For both groups, the preferred strategy to fill pauses is repetition. Where they differ is that monolingual speakers tend to fill their pauses with phrasal fillers, such as *como mencionaba* 'as I was saying', that vary from speaker to speaker. Chicano speakers, on the other hand, tend to employ more unfilled pauses and use self-correction almost twice as much as the monolinguals. What is clear is that both groups use a common set of fillers but also use different strategies to fill pauses. As far as the number of fillers is concerned, English-dominant Chicano speakers do not reduce the number of fillers in Spanish when compared to monolingual speakers. Rather, Chicano speakers use more silent pauses and add English fillers to their repertoire of Spanish fillers (in particular *well*, *like* and *so*). It is worth noting that these findings reflect the speech in an academic register of both Chicano and monolingual speakers of comparable ages and education in a university classroom setting.

Considering that a common characteristic of Spanish in the U.S. is the presence of English fillers, various studies (among them Torres & Potowski 2008, Lipski 2005, Aaron 2004, Silva-Corvalán 1994) identify the English discourse marker *so* as the most common in U.S. Spanish. In a comparative study of bilingual discourse markers in Mexican, Puerto Rican, and MexiRican speakers (one Mexican and one Puerto Rican parent) of Spanish in Chicago, Torres & Potowski (2008) find an increased use of the English discourse marker *so* for Spanish *entonces* particularly in Puerto Rican and

MexiRican speakers, suggesting that some populations are more prone to borrowing than others and that some varieties of U.S. Spanish are more receptive to English fillers than others. Proficiency, rather than generation in Torres & Potowski (2008), is the stronger predictor of the maintenance of *entonces* in favor of *so*. Previous work by Torres (2002) on Puerto Rican Spanish in New York alone had revealed proficiency in Spanish to be a salient factor on the retention of Spanish discourse markers in a contact situation with English. All generations in Torres' (2002) study used Spanish discourse markers, but the less proficient speakers of Spanish used *so* more frequently.

Lipski (2005) also finds that *so* is prevalent among heritage and immigrant speakers as well as second-language learners of Spanish in a study of speakers of Spanish across the U.S., with varying proficiencies and backgrounds. In all observed cases, *so* functions as a coordinating conjunction when used phrase-internally, and it is also attested phrase-initially and phrase-finally. Lipski points out that the range of meanings that *so* can convey is equivalent to the Spanish discourse markers *pues, así que, de manera que,* and *de modo que* among others. This wide variety of uses, along with it being a monosyllabic word, may account for its easy transfer into Spanish among bilinguals. What Lipski's study underscores is the fact that, among speakers of U.S. Spanish, *so* is used differently and can have on different functions.

Aaron (2004) finds that *so* is very widespread and frequent among bilingual speakers of Spanish born in New Mexico. Aaron compares *so* with Spanish *entonces* and finds that they have the same functions (e.g., introducing a result, marking discourse progression, etc.) and are thus in free variation. For this variety of Spanish, *so* is more

popular than *entonces*. As to the question of whether the prevalence of *so* means that *entonces* decreases in frequency, *entonces* has an added temporal function that *so* lacks. For this reason, *entonces* remains stable in bilingual New Mexican Spanish and thus *so* has no advantage over *entonces*. An added observation is that *so* may function as a trigger for code-switching in bilingual speakers.

Silva-Corvalán (1994) notes that *so* is a feature of Los Angeles Spanish present in the speech of bilingual speakers across all levels of proficiencies and even Spanish-dominant speakers, and that its use is that of a loan that replaces Spanish *así que*.

6.4 Research Questions

Based on previous findings, the research questions that guide this study are:

1. Is there a core set of fillers in the San Diego-Tijuana border area that is common to all speakers, similar to what Valdés and Geoffrion-Vinci (1998) find? Considering that speakers in their two populations used different strategies to fill pauses beyond the core fillers, are there fillers beyond the core group in San Diego-Tijuana that emerge as significant for different populations?

2. Taking into account core fillers for all speakers in addition to the population-specific fillers, how many different sets of fillers are there in the San Diego-Tijuana border area? And how are these sets different? I define a set of fillers as the inventory of fillers in a population based on number of speakers who use them and their frequency of use. This inventory of fillers consists of the core fillers for all speakers in the border area and the fillers particular to each population.

3. Since most studies cite the presence of English fillers in the Spanish of U.S.-born Hispanics, what are the English fillers that are present in the border area? Are any English fillers used significantly/consistently by any group(s)?

4. What is the status of the English filler *so* in the border area? Is it used by both groups, heritage speakers and adult immigrant speakers to the United States as Silva-Corvalán (1994) finds in Los Angeles Spanish? If so, is it used differently and with different functions across generations in San Diego, as Lipski (2005) and Aaron (2004) find in their studies? Following Torres & Potowski (2008), does an increased use of *so* result in a reduced use of *entonces*? Or is the status of *so* in the border area similar to what Aaron (2004) found for New Mexico where *so* and *entonces* are in free variation and equally stable?

This study sheds new light on the phenomenon of fillers in a Spanish-English language contact situation in the United States. It differs from previous studies in the following ways: (1) it presents data from subjects from both sides of the United States – Mexico border in the San Diego-Tijuana area. The two populations in San Diego are immigrant parents and their U.S.-born children whereas the Tijuana population represents the monolingual control group. (2) The data was gathered in an informal, mostly autobiographical interactional occasion (as defined by Erickson & Schultz 1981) and speakers used a non-academic register. (3) Thirteen different fillers were analyzed for their frequency of use and the number of speakers who use each.

6.5 Methodology

6.5.1 Normalization of the Data

Due to the varying length in interviews, the analysis was focused to 1235 words per speaker. This number corresponds to the total word count for the shortest interview (a member of the SD1 group). For all interviews, the last 1235 words in the transcription were chosen for analysis for two reasons: (1) it is toward the end of the interview that speakers felt more at ease and where the longer uninterrupted segments of speech are found; and (2), the last set of questions ask that speakers speculate about how their lives would be different on the other side of the border and what changes they envision for future generations, which are likely to be topics they have not thought about and create a more authentic conversational interaction. In contrast, the first part of the interview focuses on demographic information and language background, as previously mentioned, and thus the interaction is not as conversational.

Each transcribed segment of approximately 1235 words was coded by the researcher for all filled pauses. Unfilled pauses were not coded nor counted as they are outside of the scope of the project. The number of occurrences for each token per speaker was then counted and entered into an Excel spreadsheet.

6.5.2 Repertoire of Fillers

The 60 fillers in Table 4.1 were observed/attested in the data.

Table 6.1 All Fillers Attested in the Tijuana – San Diego Area

1	a ver	21	de hecho	41	oh
2	ah	22	digamos	*42*	*okay*
3	ahí	23	digo (que)	43	pero
4	ahora sí	24	eh	44	pienso que
5	ajá	25	¿entiendes?	45	por ejemplo
6	algo	26	entonces	46	pues
7	am	27	es que	47	que diga uno
8	así	28	eso sí	48	se que
9	así como	29	este	49	según
10	ay	30	fíjate	50	sí
11	bueno	31	haz de cuenta	51	*so*
12	como	32	igual	52	todo eso
13	como dicen	33	inclusive	53	uh
14	¿cómo le digo?	34	*like*	54	últimamente
15	¿como le dijera?	35	más o menos	55	um
16	como que	36	mm hm	56	uy
17	como quien dice	37	mmm	57	¿verdad?
18	¿cómo se dice?	38	no sé	58	y luego
19	¿cómo se puede decir?	39	¿no?	59	y todo
20	creo que	40	o sea	*60*	*you know*

Of these 60 fillers, thirteen were selected for analysis. The criteria for selection were the following:

1. Include Spanish fillers that are used by no fewer than 11 speakers (11 out of 44 total, which comprises 25% of the subjects). This limits the number of fillers under analysis to the most commonly used ones.
2. Include only Spanish fillers that are used by at least one speaker per group, taking into account both geographic and generational factors. This means limiting the number of Spanish fillers to those than can be analyzed cross-generationally and cross-geographically to determine the different sets on

both sides of the border. A Spanish filler completely absent from any of the groups is not informative in determining the set of fillers for the border area.

3. Include English fillers that appear in more than one group to determine if they are used differently depending on the language contact situation.

In Table 6.2 below I present the 13 fillers that meet the selection criteria. Eleven of these fillers are in Spanish while two (*so* and *okay*) are in English.

Table 6.2 Fillers by Numbers of Speakers

Fillers	**Groups**				
	TJ1	TJ2	SD1	SD2	All Speakers
pues	11	11	11	11	44
entonces	9	9	11	7	36
este	11	10	9	6	36
ah	5	8	9	10	32
o sea	6	10	7	7	30
bueno	9	10	7	2	28
eh	9	6	6	5	26
como que	5	8	4	7	24
digo	5	4	6	5	20
¿no?	8	7	3	1	19
¿verdad?	4	3	9	1	17
okay	3	5	2	4	14
so	0	0	3	5	8

6.5.3 Analysis

To answer the research questions, I resort to two types of analyses. First, quantitative analysis on the number of speakers who use each filler and the frequency of use for each filler will elucidate the different sets of fillers in the San Diego-Tijuana

border area. Second, qualitative analysis will determine the status of *so* in the border area.

6.6 Quantitative Analysis: Methodology, Results, and Discussion

The first set of research questions propose a description of the different sets of fillers in the border area. I define a set of fillers as one that takes into account the number of speakers and the frequency of use for each filler. As such, in section 6.6.1 I analyze the sets of fillers based on the number of speakers who use each filler and in section 6.6.2 I analyze the sets of fillers based on the frequency of use for each filler. Finally, in section 6.6.3 I provide a discussion of the results of the quantitative analysis.

6.6.1 Number of Speakers

For each of the four groups that are part of the study (TJ1, TJ2, SD1, and SD2), I divide the thirteen fillers in four tiers that indicate the number of speakers for each. The fillers in Tier 1 are used by no fewer than nine speakers in a group (of 11 speakers total). The fillers in Tier 1 are the ones that most speakers use. The fillers in Tier 2 are attested in between six and eight speakers per group. The fillers in Tier 3 are attested in between three and five speakers, and the fillers in Tier 4 in only one or two speakers. This last tier contains the fillers that the smallest number of speakers uses. These results are summarized in Table 6.3.

Table 6.3 Fillers by Number of Speakers – Four Tiers

Group	Tier 1 (9-11 speakers use the filler)	Tier 2 (6-8 speakers use the filler)	Tier 3 (3-5 speakers use the filler)	Tier 4 (1-2 speakers use the filler)
TJ1	Pues (11) Este (11) Entonces (9) Bueno (9) Eh (9)	¿No? (8) O sea (6)	Ah (5) Como que (5) Digo (5) ¿Verdad? (4) Okay (3)	
TJ2	Pues (11) Este (10) Bueno (10) O sea (10) Entonces (9)	Ah (8) Como que (8) ¿No? (7) Eh (6)	Okay (5) Digo (4) ¿Verdad? (3)	
SD1	Pues (11) Entonces (11) Este (9) Ah (9) ¿Verdad? (9)	Bueno (7) O sea (7) Eh (6) Digo (6)	Como que (4) ¿No? (3) So (3)	Okay (2)
SD2	Pues (11) Ah (10)	Entonces (7) Como que (7) O sea (7) Este (6)	Eh (5) Digo (5) So (5) Okay (4)	Bueno (2) ¿Verdad? (1) ¿No? (1)

A cursory analysis of the four tiers yields the following observations about the number of speakers per filler in the border area:

1. Both Tijuana groups as well as the San Diego first-generation group have five fillers in Tier 1. This is very different from what happens with the San Diego second-generation speakers, for whom only two fillers are in Tier 1 and therefore attested in the majority of speakers.
2. The only English filler attested in all four groups is *okay*, but it never places higher than Tier 3 (five speakers).

3. The English filler *so* is attested in both San Diego groups in Tier 3. However, it is more prevalent among speakers in the Heritage group (SD2).

This indicates that while English fillers have a presence in all Spanish-speaking populations on both sides of the border, only one (*okay*) is attested in Tijuana and two (*okay* and *so*) in both San Diego groups.

6.6.1.1 Quantitative Analysis based on the Number of Speakers

For each of the four groups I determined the number of speakers for whom each filler is attested at least once in the data. These numbers are reported in Table 6.4. I have also organized the data into the following groupings to test for various effects:

1. To test for a geographic effect, I combine all 22 San Diego speakers (SD1 and SD2) and compare them with all 22 Tijuana speakers (TJ1 and TJ2)
2. To test for a heritage effect, I combine all 33 non-heritage speakers (SD1, TJ1 and TJ2) and compare them with the 11 heritage speakers (SD2)
3. To test for an immigration effect, I compare the 11 speakers in the immigrant group (SD1) with the 22 monolingual Tijuana controls (all TJ).

Table 6.4 Number of Speakers who Use the Filler at Least Once per Population & Groupings for Comparisons

	ah	bueno	como que	digo	eh	entonces	este	¿no?	o sea	okay	pues	¿verdad?	so
Individual populations (11 speakers each group)													
TJ1	5	9	5	5	9	9	11	8	6	3	11	4	0
TJ2	8	10	8	4	6	9	10	7	10	5	11	3	0
SD1	9	7	4	6	6	11	9	3	7	2	11	9	3
SD2	10	2	7	5	5	7	6	1	7	4	11	1	5
Geographical comparison (22 speakers each group)													
All TJ	13	19	13	9	15	18	21	15	16	8	22	7	0
All SD	19	9	11	11	11	18	15	4	14	6	22	10	8
Heritage comparison (33 non-Heritage speakers, 11 Heritage speakers)													
Non-Heritage	22	26	17	15	21	29	30	18	23	10	33	16	3
Heritage	10	2	7	5	5	7	6	1	7	4	11	1	5
All (44 speakers total)													
All speakers	32	28	24	20	26	36	36	19	30	14	44	17	8

For each filler I create a 2x2 contingency table that compares the number of speakers who use a given filler with the number of speakers who do not. This is done for the two populations that are being compared. In Table 6.5 I provide the contingency table to test for a geographical effect on the number of speakers who use *ah* in Tijuana and San Diego.

Table 6.5 2x2 Contingency Table for Filler *ah* in Tijuana and San Diego

	Number of speakers who use *ah*	Number of speakers who don't use *ah*	Totals
Tijuana	13	9	22
San Diego	19	2	22
			44

A Pearson's chi-square test of association would normally be calculated to test for the significance between the differences of the two populations. However, the chi-square test would not only be unreliable given the small sample size (22 speakers per group, 44 total), but it would also be impossible since not all of the expected frequencies are greater than or equal to 5 (as evidenced by the fact that the number of speaker who do not use *ah* in San Diego is only 2). The non-parametric Fisher Exact Probability Test is used instead since it is better-suited for the analysis of contingency tables whose sample size is small and with expected cell frequencies less than 5. The null hypothesis in this case is that there is no statistical difference in the number of speakers who use a filler (for example *ah*) between two populations being compared (TJ and SD, for example). Analyzing the contingency table for use of *ah* between Tijuana and San Diego for number of speakers using the Fisher Exact Probability Test results in a p-value of 0.044, which is statistically significant. This result is presented in Table 6.6.

Table 6.6 Fisher Exact Probability Test for Number of Speakers who use *ah* in Tijuana and San Diego

	Number of speakers who use *ah*	Number of speakers who don't use *ah*	Totals
Tijuana	13	9	22
San Diego	19	2	22
			44
Fisher Exact Probability Test			p = 0.044

Because the number of speakers who use this filler in San Diego is 19 compared to 13 in Tijuana, I conclude that the filler *ah* is statistically significant for a greater number of users in San Diego.

In Table 6.7 I present the results of the Fisher Exact Probability Test for the three comparisons: geographical, heritage, and immigrant. Notice, for example, that in the case of the filler *bueno* two comparisons (geographical and heritage) yield statistically significant results (p=0.002 and p=0.0005). The lower of these two is the most significant one (i.e., heritage comparison p=0.0005). Similarly, in the two cases where all three comparisons yield statistically significant results (i.e., *¿no?* and *so*), it is the lowest p-value of the three that I consider the most significant.

Table 6.7 Results of the Fisher Exact Probability Test. The most significant p-values are accompanied by a label in parenthesis to indicate the population for whom the statistical effect is most significant

	Geographical	**Heritage**	**Immigrant**
	TJ vs SD	Non-Heritage vs Heritage	All TJ vs SD1 Immigrants
ah	0.044 (San Diego)	0.117	0.181
bueno	0.002	0.0005 (Non-Heritage)	0.146
como que	0.381	0.365	0.194
digo	0.381	0.637	0.354
eh	0.178	0.238	0.347
entonces	0.65	0.091	0.178
este	0.022	0.015 (Non-Heritage)	0.25
¿no?	0.0009 (Tijuana)	0.008	0.039
o sea	0.373	0.49	0.439
okay	0.373	0.49	0.256
pues	1	1	1
¿verdad?	0.268	0.02	0.008 (SD1 Immigrants)
so	0.001 (San Diego)	0.015	0.03

There are seven fillers for which the Fisher Exact Probability Test yields no statistical significance for any of the three comparisons (geographical, heritage, and immigrant). This means that as far as number of speakers is concerned, seven fillers are not used

significantly by more speakers in any one group over any other group. As such, these seven fillers are the core fillers based on number of speakers for the San Diego-Tijuana border area. They are presented in (2).

(2) Core fillers in the San Diego-Tijuana border area
(based on number of speakers)
como que, *digo*, *eh*, *entonces*, *o sea*, *okay*, and *pues*

Based on the results from Table 6.7, the fillers that are statistically significant for the different populations can be summarized as follows in Table 6.8.

Table 6.8 Statistically-significant Fillers per Population based on Number of Speakers

Core Fillers (all groups)	San Diego (SD1 & SD2)	Tijuana (TJ1 & TJ2)	Non-Heritage (TJ1, TJ2, SD1)	Immigrants (SD1)
como que digo eh entonces o sea okay pues	ah so	¿no?	bueno este	¿verdad?

This analysis on the number of speakers who activate each filler, and which fillers are significant for what population, is not a complete picture of what the sets of fillers are in the border area as it merely provides a general idea of what the core fillers might be and which populations of speakers are activating what fillers more than others. What is also necessary is an analysis of the frequency of use for each filler per population. I provide this analysis in the following section.

6.6.2 Frequency of Use

To determine the frequency of use for each filler, I take the number of times each speaker uses each filler and find the median number of instances of use for each group (SD1, SD2, TJ1, and TJ2). The median deals with outliers and handles high within-group variation to describe frequency of use more accurately than the mean for the group. Consider for instance Table 6.9, which presents examples of the number of tokens per speaker for three fillers (¿*verdad*?, *ah*, and *digo*) for some groups in the San Diego population and compares their medians and means per group.

Table 6.9 Examples of Number of Tokens per Speaker

Filler	Group	Tokens per speaker	Median	Mean
¿verdad?	SD1	0-0-1-1-1-**2**-2-3-4-5-20	2	3.54
ah	SD2	0-1-1-3-4-**5**-6-7-13-40	5	7.27
digo	SD2	0-0-0-0-0-**0**-1-1-3-3-17	0	2.27

The mean for ¿*verdad*? and *ah* for the SD1 and SD2 groups, respectively, are driven by the high occurrence of the filler in one speaker for each group (20 instances of ¿*verdad*? for one speaker in SD1 and 40 instances of *ah* for one speaker in SD2). As such, the means are not truly representative of the use of the filler by each of the groups. This is demonstrated by the use of *digo* in the SD2 population where the mean of 2.27 is not only driven by the one speaker who uses it 17 times in sample but it also does not reflect the frequency of use by the group since more than half of the speakers in SD2 do not use this filler. A median of 0 more accurately represents the frequency of use of *digo* in SD2 than the mean does. Furthermore, it is worth noting that a median of 0 does not

mean zero use in a given group. Rather, it means that the filler is very low frequency for the group.

For each of the four groups that are part of the study (TJ1, TJ2, SD1, and SD2), I divide the thirteen fillers in four tiers that indicate the frequency of use for each. Just as I presented the tiers in section 6.6.1 for the number of speakers, the fillers in Tier 1 are the most frequently used with a median of 10 of higher for the group. The fillers in Tier 1 are the most frequently used. The fillers in Tier 2 have a median between four and nine. The fillers in Tier 3 have a median between one and three, and the fillers in Tier 4 have a median of zero. This last tier contains the least frequently used fillers for the group. These results are summarized in Table 6.10.

Table 6.10 Fillers by Frequency of Use – Four Tiers

Population	Tier 1 (Median: 10+)	Tier 2 (Median: 4-9)	Tier 3 (Median: 1-3)	Tier 4 (Median: 0)
TJ1	Pues (14) Este (13)	Entonces (6)	No? (3) Bueno (2) Eh (2) O sea (2)	Ah (0) Como que (0) digo (0) okay (0) verdad? (0)
TJ2	Pues (26)	Este (9) Entonces (7) O sea (4)	No? (2) Bueno (2) Eh (1) Ah (1) Como que (1)	Digo (0) okay (0) verdad? (0)
SD1	Pues (18)	Este (5) Entonces (5)	Ah (2) Verdad? (2) bueno (1) eh (1) o sea (1) digo (1)	No? (0) Como que (0) okay (0) so (0)
SD2	Pues (10)	Ah (5)	entonces (2) este (1) o sea (1) como que (1)	Verdad? (0) bueno (0) eh (0) digo (0) no? (0) okay (0) so (0)

The following conclusions can be drawn with an overview of the four tiers based on frequency of use of fillers in the border area:

1. 1. *Pues* is the most popular filler in the border area for all groups. Its highest frequency is in the TJ2 group and its lowest in the SD2 group.
2. It appears that overall the SD2 group uses fillers with less frequency than any of the other groups since 7 out of 13 fillers have a frequency median of zero, and three more have a frequency median of 1.

3. All of the English fillers for all groups have a median of 0. They are very low frequency even in the San Diego populations where more speakers use them.
4. The only English filler attested in all four groups is *okay* but its frequency is very low.

These observations suggest that heritage speakers use fillers less often than any other group. Because the use of silent pauses and repetitions is outside the scope of this study, it is possible that heritage speakers may reduce their fillers and instead use silent pauses and repetitions. This question remains as a possible avenue for further study. Moreover, all English fillers in the border area have a frequency median of 0. This may suggest that speakers add English fillers to their repertoire but these English fillers are not affecting the frequency of Spanish fillers.

6.6.2.1 Quantitative Analysis based on Frequency of Use

For each of the four groups I computed the number of times that each filler is used by each speaker. I then determined the median for each of the four groups. These medians are reported in Table 6.11. I organized the data into the same groupings as I did in section 6.6.1.1 to test for various effects. These groupings and effects are the following:

1. To test for a geographic effect, I combine all SD (SD1 and SD2) and compare them with all TJ (TJ1 and TJ2).

2. To test for a heritage effect, I combine all non-heritage speakers (SD1, TJ1 and TJ2) and compare them with the heritage speakers (SD2).
3. To test for an immigration effect, I compare the immigrant group (SD1) with the monolingual Tijuana controls (all TJ).

Table 6.11 Medians per Population & Groupings for Comparisons

	ah	bueno	como que	digo	eh	entonces	este	no?	o sea	okay	pues	¿verdad?	so
TJ1	0	2	0	0	2	6	13	3	2	0	14	0	0
TJ2	1	2	1	0	1	7	9	2	4	0	26	0	0
SD1	2	1	0	1	1	5	5	0	1	0	18	2	0
SD2	5	0	1	0	0	2	1	0	1	0	10	0	0
All TJ	1	2	1	0	2	6.5	10.5	2.5	2	0	19.5	0	0
All SD	3.5	0	0.5	0.5	0.5	3	2.5	0	1	0	12.5	0	0
Non-Heritage	1	1	1	0	2	6	8	1	2	0	18	0	0
Heritage	5	0	1	0	0	2	1	0	1	0	10	0	0
Overall median for all speakers	1.5	1	1	0	1	5	7.5	0	2	0	17	0	0

For the three comparisons (geographical, heritage, immigration) for statistical significance, I use Mood's Median test. This test is used to compare the medians of two groups: one group consists of all frequencies that are equal or below the overall median for all speakers, and another group of frequencies that are higher than the overall median. The null hypothesis in this case establishes that the medians of the two groups are not significantly different. Consider, for example, the comparison for the filler *ah* between Tijuana and San Diego. In Table 6.11 above (Medians per population & groupings for

comparisons), the overall median for all speakers for *ah* is 1.5. For a comparison of the medians between Tijuana and San Diego I determine the number of speakers (out of 22) for each population who are lower or equal to the overall median of 1.5 and the number of speakers who are higher than 1.5. This is presented in Table 6.12 below.

Table 6.12 Median Comparison for *ah* – Tijuana and San Diego

Medians for *ah* in Tijuana per speaker, 22 speakers total 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 1 - 1 - 1 - 1 - 1 - 1 - 2 - 2 - 2 - 3 - 5 - 6 - 7	
Number of speakers in Tijuana lower or equal to the overall median of 1.5	15
Number of speakers in Tijuana higher than the overall median of 1.5	7
Medians for *ah* in San Diego per speaker, 22 speakers total: 0 - 0 - 0 - 1 - 1 - 1 - 1 - 2 - 2 - 3 - 3 - 4 - 4 - 5 - 5 - 6 - 7 - 11 - 13 - 17 - 22 - 40	
Number of speakers in San Diego lower or equal to the overall median of 1.5	7
Number of speakers in San Diego higher than the overall median of 1.5	15

In Table 6.13 I provide the 2x2 contingency table for the number of speakers whose frequency use of *ah* is lower or equal to the overall median, and those whose use of the same filler is higher than the overall median. Given the small sample size (44 speakers total) and the fact that the expected cell frequencies will contain values less than 5, the Fisher Exact Probability Test is used to test for a geographical effect for the frequency of use of the filler *ah* in Tijuana and San Diego using the Fisher Exact Probability Test.

Table 6.13 2x2 Contingency Table and Fisher Exact Probability Test for *ah* San Diego vs Tijuana

Overall median of *ah* for all speakers: 1.5			
	Number of speakers lower or equal to the overall median	Number of speakers higher than the overall median	Totals
Tijuana	15	7	22
San Diego	7	15	22
			44
Fisher Exact Probability Test			p = 0.016

With a p-value of 0.016, the difference in the frequency of use if *ah* in San Diego compared to Tijuana is statistically significant. Because the number of speakers who are higher than the median in San Diego is 15, compared to 7 in Tijuana, the filler *ah* is statistically significant for greater frequency among speakers in San Diego.

In Table 6.14 I present the results of the Mood's Median Test which uses Fisher's Exact Probability Test for the three comparisons: geographical, heritage, and immigrant. As it was the case with the comparisons in section 6.6.1 for number of speakers, some of the comparisons on this table for frequency of use may yield more than one result with statistical significance. For example, *¿no?* is statistically significant for all three comparisons but the most significant is the lowest of these three (i.e., geographical comparison, p=0.0009). In Table 6.14 below, the most statistically significant p-values are accompanied by a label in parenthesis to indicate the population for whom the statistical effect is most significant.

Table 6.14 Results of Mood's Median Test for Frequency of Use

	Geographical	Heritage	Immigrant
	TJ vs SD	Non-Heritage vs Heritage	all TJ vs SD1
ah	0.016 (San Diego)	0.081	0.085
bueno	0.062	0.013 (Non-Heritage)	0.45
como que	0.262	0.564	0.181
digo	0.381	0.634	0.354
eh	0.182	0.147	0.571
entonces	0.016	0.008 (Non-Heritage)	0.357
este	0.003 (Tijuana)	0.243	0.009
no?	0.0009 (Tijuana)	0.008	0.029
o sea	0.051 (Tijuana)	0.061	0.267
okay	0.373	0.49	0.256
pues	0.183	0.081	0.269
verdad	0.268	0.02	0.008 (Immigrants)
so	0.001 (San Diego)	0.015	0.03

There are five fillers for which Mood's Median Test yields no statistical significance for any of the three comparisons (geographical, heritage, and immigrant). In regards to frequency of use, five fillers are not used more frequently by any one group over any other group. These five core fillers based on frequency of use are presented in (3).

(3) Core fillers in the San Diego-Tijuana border area
(based on frequency of use)
como que, *digo*, *eh*, *okay*, and *pues*

A closer look at the overall median for all speakers for these five core fillers (presented above in Table 6.11: Medians per population & groupings for comparisons) reveals that *pues* is the only core filler with high frequency (median of 17). These medians are repeated below in abbreviated form in Table 6.15. The low overall medians for the other

core fillers (*como que*, *digo*, *eh*, and *okay*) suggest that the lack of statistical significance is due to their infrequency across all populations.

Table 6.15 Overall Median for all Speakers for all Fillers

	ah	bueno	como que	digo	eh	entonces	este	no?	o sea	okay	pues	¿verdad?	so
Overall median for all speakers	1.5	1	1	0	1	5	7.5	0	2	0	17	0	0

Based on the results from Table 6.15, the fillers that are statistically significant for the different populations can be summarized as follows in Table 6.16. I provide in parenthesis after every filler its overall median for all speakers to indicate whether the filler in question is frequent or infrequent in the border area.

Table 6.16 Statistically-significant Fillers per Population Based on Frequency of Use

Core Fillers (all groups)	San Diego (SD1 & SD2)	Tijuana (TJ1 & TJ2)	Non-Heritage (TJ1, TJ2, SD1)	Immigrants (SD1)
como que (1) digo (0) eh (1) okay (0) pues (17)	ah (1.5) so (0)	¿no? (0) este (7.5) o sea (2)	bueno (1) entonces (5)	¿verdad? (0)

Of the thirteen fillers analyzed, eight are considered low frequency with an overall median of 0 or 1. Section 6.6.1 revealed that eight out of thirteen fillers are activated by more than half of the speakers. However, this section (6.6.2) reveals that though

speakers may be activating those fillers, only five of thirteen fillers are being used with medium or high frequency.

6.6.3 Discussion of Quantitative Analysis

In this section I provide a comparison of the two quantitative analyses in order to determine the number of different sets of fillers in the border area and to provide a description of each. With respect to the core fillers, the statistical analysis for number of speakers and frequency of use match five of the fillers. These five fillers – shown in (4) – are thus the core fillers for the San Diego-Tijuana area.

(4) Core fillers in the San Diego-Tijuana border area
como que, *digo*, *eh*, *okay*, and *pues*

The geographical comparison for both number of speakers and frequency of use (Table 6.17 below) also provides a match for three of the fillers (*ah* and *so* as statistically significant for San Diego and *¿no?* for Tijuana). Both the analysis on number of speakers and frequency of use coincide in finding *ah* and *so* statistically significant for San Diego, but the analysis on frequency of use describes the Tijuana fillers more fully in that in addition to *no?* being statistically significant for Tijuana, *este* and *o sea* also show statistically significance for Tijuana based on frequency of use (but not number of speakers alone). This difference underscores the importance of taking into account both number of speakers and frequency of use. In the analysis for number of speakers, *este* emerges as statistically significant for non-heritage speakers (a group that includes Tijuana speakers) and *o sea* as a core filler for all groups. However, though more non-heritage speakers activate *este* than heritage speakers, Tijuana speakers use it more

frequently than San Diego speakers. The case is somewhat similar for *o sea*, which is all speakers activate (hence its status as a core filler based on number of speakers) but Tijuana speakers use it more frequently than San Diego speakers.

Table 6.17 Geographical Comparison: Number of Speakers and Frequency of Use

Geographical Comparison (TJ vs SD)		
	Number of Speakers	Frequency of use
ah	0.044 (San Diego)	0.016 (San Diego)
bueno	0.002	0.062
como que	0.381	0.262
digo	0.381	0.381
eh	0.178	0.182
entonces	0.65	0.016
este	0.022 (Non-heritage)	0.003 (Tijuana)
¿no?	0.0003 (Tijuana)	0.0009 (Tijuana)
o sea	0.373 (core)	0.051 (Tijuana)
okay	0.373	0.373
pues	1	0.183
¿verdad?	0.268	0.268
so	0.001 (San Diego)	0.001 (San Diego)

According to the geographical comparison, the fillers for San Diego and Tijuana are the following:

(5) Fillers for Tijuana and San Diego: Geographical Comparison
Tijuana: *este*, *no?*, *o sea*
San Diego: *ah*, *so*

The heritage comparison (Table 6.18) reveals that *bueno* is statistically significant for non-heritage speakers both based on number of speakers and frequency of use. However, *entonces*, a core filler based on number of speakers, is more frequently

used by non-heritage speakers than heritage speakers. I discussed the status of *este* as a Tijuana filler based on frequency of use in the previous paragraph.

Table 6.18 Heritage Comparison: Number of Speakers and Frequency of Use

Heritage Comparison (All TJ + SD1 vs SD2)		
	Number of Speakers	Frequency of use
ah	0.117	0.081
bueno	0.0005 (Non-heritage)	0.013 (Non-heritage)
como que	0.365	0.564
digo	0.637	0.634
eh	0.238	0.147
entonces	0.091 (core)	0.008 (Non-heritage)
este	0.015 (Non-heritage)	0.243 (Tijuana)
¿no?	0.008	0.008
o sea	0.49	0.061
okay	0.49	0.49
pues	1	0.081
¿verdad?	0.02	0.02
so	0.015	0.015

No fillers emerge as significant for heritage speakers neither by number of speakers nor by frequency of use. This may suggest that heritage speakers in the border area complement their sets of fillers with English fillers that are outside the scope of this study. Recall from section 6.5 that one of the criteria for including English fillers in the study is that they must manifest in more than one group. The only two English fillers to do so are *okay* and *so*. However, two English fillers (presented in section 6.5.3) that are unique to the heritage group are *like* and *you know*. Therefore, these two English fillers can be said to be a part of the heritage set of fillers but they are outside of the scope of

this study. The lack of statistical significance of any Spanish fillers for the heritage group, coupled with the presence of English fillers unique to this population, may indicate that the frequency of fillers is distributed between Spanish and English fillers, since this group has access to both sets, and thus no filler emerges with a statistically significant difference for this group. There is also the possibility that heritage speakers use more silent pauses and repetitions, which were not considered in this study.

It is worth bringing up the English filler *so* at this point in the discussion. One would expect it to be more statistically significant for the heritage group than for the San Diego group, since the heritage group is the most English dominant group in the border area. Nonetheless, with a median of 0 for both San Diego populations, *so* is equally infrequent for both. This may indicate that in an informal register in Spanish in the border, English fillers are first attested in the most English-dominant group (SD2) with low frequency (for example *like*, *you know*, and *so*), then one of these English fillers (e.g., *so*) surfaces in the group that has the most interaction with the English-dominant group (SD1) in equally-low frequency, and over time an English filler such as *okay* may surface in enough speakers in English-dominant and Spanish-monolingual populations in low frequency. For a more detailed discussion on the functions of *okay* and *so* in the border area, see section 6.7 of this chapter.

According to the heritage comparison, the fillers for heritage and non-heritage are the following (6):

(6) Fillers for Non-heritage and Heritage Speakers: Heritage Comparison
Non-heritage: *bueno*, *entonces*
Heritage: English fillers such as *like* and *you know*

Finally, for the immigrant comparison (Table 6.19) both the analysis on number of speakers and frequency of use reveal that *verdad?* is statistically significant for the immigrant (SD1) group.

Table 6.19 Immigrant Comparison: Number of Speakers and Frequency of Use

Immigrant Comparison (All TJ vs SD1)		
	Number of Speakers	Frequency of use
ah	0.181	0.085
bueno	0.146	0.45
como que	0.194	0.181
digo	0.354	0.354
eh	0.347	0.571
entonces	0.178	0.357
este	0.25	0.009
no?	0.039	0.029
o sea	0.439	0.267
okay	0.256	0.256
pues	1	0.269
verdad	0.008 (SD1)	0.008 (SD1)
so	0.03	0.03

Based on the two analyses (number of speakers and frequency of use) and the three comparisons (geographical, heritage, and immigrant) I conclude that there are three sets of fillers in the border area: Tijuana, Immigrant, and Heritage. These three sets are exemplified in Table 6.20 below followed by a list of fillers that belong to each category in Table 6.21. The fillers presented in Table 6.21 are accompanied by the number of speakers (out of 44 in total) and its overall median for frequency of use in parenthesis.

Table 6.20 Three Sets of Fillers in the Border Area

Set	Types of fillers			
Tijuana	Core fillers	Tijuana fillers	Non-heritage fillers	
Immigrant	Core fillers	San Diego fillers	Non-heritage fillers	Immigrant fillers
Heritage	Core fillers	San Diego fillers	Other English fillers	

Table 6.21 Fillers in Each Category with Number of Speakers and Overall Median

	Number of speakers n=	%	Frequency of use (overall median)
Core Fillers			
como que	24	55%	1
digo	20	45%	0
eh	26	59%	1
okay	14	32%	0
pues	44	100%	17
Tijuana			
este	36	82%	7.5
no?	19	43%	0
o sea	30	68%	2
San Diego			
ah	32	73%	1.5
so	8	18%	0
Non-Heritage			
bueno	28	64%	1
entonces	36	82%	5
Immigrant			
verdad?	17	39%	0

A correlation to determine the strength and the direction of a relationship between the number of speakers and the overall frequency of use results in a correlation coefficient *r* of 0.78677. This points to a strong relationship between the number of speakers and frequency of use, and it indicates that as more speakers activate a filler

said filler increases in overall frequency. Table 6.21 also reveals that only three fillers occur in high frequency in the border area (overall median > 2) and the rest are evenly distributed between medium (overall median of 1-2) and low frequency (overall median <1). Looking at Table 6.21, it appears as though, except for San Diego and heritage speakers, one filler dominates in frequency per comparison: for the core fillers it is *pues*; for Tijuana it is *este*; and for non-heritage it is *entonces*.

6.7 Qualitative Analysis: English fillers in the Spanish in the Border Area

The two English fillers in the Spanish of the border area are *okay* and *so*. The quantitative analysis revealed that *okay* is a core filler for all populations whereas *so* is only statistically significant for the San Diego groups and thus a filler in the immigrant and heritage sets of fillers . I have established that both of these English fillers are low frequency (overall median for frequency of use = 0). In this section I provide a qualitative analysis of *okay* and *so* to investigate whether their functions and uses are the same across the different populations that use them.

6.7.1 The Status of *okay* in San Diego and Tijuana

Not many studies have focused on *okay* as a discourse marker in Spanish-English bilingual speech. A study by Vickers & Goble (2011) on the use of discourse markers among Spanish-dominant and Spanish-English bilinguals during medical visits in California reveals that *okay* occurs in very low frequency among the Spanish-dominant speakers and slightly more frequently among Spanish-English bilinguals,

particularly those who are English-dominant. In that same study, its uses as a discourse marker in English-dominant speakers are as a topic shifter and as a floor shifter (Vickers & Goble 2011).

In the San Diego-Tijuana border area *okay* is also low-frequency (recall that its overall median for frequency of use is 0) but there is no significant difference in the number of speakers who activate on either side of the border. In Table 6.22 below I provide the distribution of *okay* per population both for number of instances and number of different speakers.

Table 6.22 Instances of *okay*

	Instances per population	Number of speakers
SD1	3	3
SD2	10	4
TJ1	10	3
TJ2	14	5

Okay is attested 37 times in the border area across all four populations. There is almost no difference in the number of speakers who use *okay* on either side of the border as the filler is used by seven speakers in San Diego and eight in Tijuana. Of these, the San Diego first-generation group (SD1) has variable contact with English and the San Diego second-generation group (SD2) can be said to be English-dominant. However, contact with English, or use of the discourse marker *okay* in English, does not appear to have an effect on the number of speakers who use it. However, for number of instances it is more frequent in Tijuana (24 tokens in total) than in San Diego (13 tokens in total), with

the San Diego first-generation group (SD1) being the one with the lowest number of instances.

The filler *okay* has four functions in the Spanish of San Diego and Tijuana as attested in the corpus. *Okay* is used:

1. As an affirmative response
2. To express approval and acceptance
3. To begin a discourse segment
4. To return to the original topic or move the current topic along

What follows are examples of the four different uses of *okay* among speakers in the border area. In (7) I present an example of *okay* as it is used by an SD1 generation speaker as an affirmative response. In (8-9) first-generation speakers in Tijuana and San Diego, respectively, use *okay* to express approval and acceptance. In (10-11) *okay* is used to begin a discourse segment as a direct response to a question. And in (12-13) *okay* is used to return to the original topic or move the current topic along.

(7) *Dicen "amá, tú hablas el inglés bien funny".*

‘They say, “Mom, you speak English really funny”.’

Digo "ah, ¿pero me entendiste?

‘I say, “ah, but you understood?”’

Y dice "okay, sí".

‘And he says, “okay, yes”.’

[SD B1]

(8) *Y me dicen ¿sabes qué? Ve por un refresco, okay, vete por una soda.*

'And they tell me, you know what? Go get a soda, okay, go get a soda.'

Y así son palabras y ellos se quedan así como que "okay, está bien" siempre.

'And those are the words and they're like "okay, that's fine" always.'

[TJ P1]

(9) *Dame la tortilla. Okay, se la das.*

'Give me the tortilla. Okay, you give it to them'.

[SD J1]

(10) *Okay sí, este, yo estaba en el- estaba en el ejército.*

'Okay, yes, well, I was in the- I was in the army.'

[TJ Q1] in direct response to the question of why the subject was living in Mexico City.

(11) *Okay pues donde vivo yo, en National City, aquí está muy calmado.*

'Okay, well, where I live, in National City, here it's very calm'.

[SD H2] in direct response to the question of what it is like to live here.

(12) *La empujó bien recio, que porque no cerraba,*

'She pushed in (the door) very hard, supposedly because it wouldn't close,'

y okay, ya ya me cambié y todo y ya quería salir.

'and okay, I changed clothes and everything and I wanted to get out.'

[SD A2]

(13) *Porque dicen "no te entiendo" y okay empiezan a hablar que mmm*

'Because they say, "I don't understand" and okay they start talking like mmm

ay ¿cómo le podré decir?

"How could I tell him/her?"'

[TJ P1]

Given the low frequency of *okay* across all populations, not all functions are attested in all four groups. In Table 6.23 I present the number of attestations of *okay* per function for each of the four populations. No definitive conclusions can be drawn given that the small amount of attestations per group is distributed among a small number of speakers who activate the filler (see Table 6.22 for number of speakers). However, it appears as though the most frequent use of *okay* in the border area is to begin a discourse segment followed by its use to express approval and acceptance. A larger sample with more conversational data is necessary to clarify this issue in a future study.

Table 6.23 Functions of *okay* per Population. Number indicates number of attestations.

	TJ1	TJ2	SD1	SD2	Total
As an affirmative response	1	2	1	2	6
To express approval and acceptance	5	2	2	0	9
To begin a discourse segment	3	10	0	7	20
To return to the original topic/ move the current topic along	1	0	0	1	2
Total	10	14	3	10	37

While it may be difficult to ascertain with a limited sample size whether the functions of *okay* are significantly different across populations, the categorization of *okay* as a filler may vary according to the language contact situation for each group. Lipski (2005) differentiates between a lexical borrowing and an English filler that is inserted into Spanish. For the Tijuana groups, *okay* follows Lipski's definition of a lexical borrowing in that in Tijuana it has become lexicalized and it is used consistently, and it is even used by Spanish monolinguals whose knowledge of English is non-existent or limited. Its status in San Diego is more difficult to determine. Based on Lipski's (2005) analysis, *okay* may represent a case of single-word codeswitching for English-dominant SD2 speakers who may retain the English phonotactics in its pronunciation. For SD1 speakers, its status may be intermediate between lexical borrowing and inserted English filler depending not only on the phonotactics but also on whether its functions or uses can be determined to be similar to its uses in English. These questions are beyond the scope of this study considering the small sample size for *okay*, but it is nonetheless important to consider the possibility of fillers being categorized according to the language contact situation.

6.7.2 The Status of *so* in San Diego and Tijuana

In section 6.3 I provided an overview of previous research on the filler *so* in Spanish-English bilinguals in the U.S. The quantitative analysis in section 6.7 revealed that while *so* is statistically significant for the San Diego populations, as it does not emerge at all in Tijuana Spanish, it is also low frequency for SD1 and SD2 speakers. In this section, I provide a qualitative analysis of *so* for both San Diego populations to determine whether the differences in language experiences between parents and children in San Diego result in different uses and functions for *so*.

Like *okay,* the filler *so* is very low frequency in the corpus. Its overall median for frequency of use is 0. Unlike *okay*, *so* is not attested in all four populations but it is rather attested only in San Diego. The quantitative analysis revealed a geographic effect and *so* is statistically significant for the San Diego group. In Table 6.24 I provide the distribution of *so* across both generations in San Diego both for number of instances and number of different speakers.

Table 6.24 Instances of *so*

	Instances per population	Number of speakers
SD1	11	3
SD2	26	5

So is attested 37 times in both generations in San Diego. It is used significantly more among second-generation San Diego speakers, both in terms of frequency of use and number of speakers. In fact, all three comparisons (geographical, heritage, and immigrant) were statistically significant for *so* in the quantitative analysis, but the

strongest effect ($p<0.001$) was a geographical one. Since the San Diego second-generation group (SD2) is English-dominant, it is this group that exhibits the greater use of *so*.

The filler *so* has four different functions attested in San Diego. These are:

1. To express cause and/or consequence
2. To change topics
3. To begin a discourse segment
4. To cede the floor[2]

Below I provide examples for the four different uses of *so* in San Diego. In (14-15) *so* is used to express cause or consequence. In (16-17) it is used to change topics. In (18-19) it is used to begin a discourse segment. And in (20-21) it is used to cede the floor.

(14) Cause Consequence

Ya tengo bastante que no voy, so este año sí, ya pedí mis vacaciones, voy a ir. Estas vacaciones que salen de la escuela, voy a ir.

'It's been a while since I don't go, so this year yes, I already requested vacation, I'm going to go. This school break, I'm going to go.'

[SD B1]

[2] The use of *so* to cede the floor indicates that a speaker can end an utterance with *so* as a way of indicating that they have finished and the other person can speak. This generally has a trailing effect on the vowel sound of *so* and can be accompanied with nonverbal cues such as eyebrow raising.

(15) Cause Consequence

Y a veces no me gusta ir cuando van ellos, so a veces nomás me quedo en la casa o hago cosas así de ese tipo.

'And sometimes I don't like to go when they go, so sometimes I just stay at home or I do things like that'.

[SD F2]

(16) Change Topics

Bueno, estoy en el departamento de alfabetización, so la gente no sabe hablar muy bien el inglés, por eso -- ayudo a la gente.

'Well, I'm in the literacy department, so people cannot speak English very well, that's why-- I help people'.

[SD C1]

(17) Change Topics

La rutina estar en la escuela es igual ¿no? So, siempre íbamos por a veces un mes, pero eso – eso era cada en cuando.

´The routine of being in school was the same, right? So we'd always go for sometimes a month but that – that was every once in a while'.

[SD B2]

(18) Begin a Discourse Segment

So es uno y uno, sí, ¿cómo se dice?, spanglish.

'So it's one and one, yes. How do you say that? Spanglish'.

[SD L2]

(19) Begin a Discourse Segment

So enseño una clase que es una aerobics class de Zumba, que es una clase de aerobics.

'So I teach a class that's an aerobics Zumba class, which an aerobics class'.

[SD M2]

(20) Cede the Floor

Más cómodo pues como por el español porque lo aprendí -- fue mi primer lenguaje, aprendí, so más cómodo puedo, you know, secretear o, you know, mucha gente no habla el español so, so, me siento más cómodo con el español. Pero, um, en inglés es importante so.

'More comfortable in Spanish because I learned it – it was my first language, I learned, so more comfortably I can, you know, gossip or, you know, a lot of people don't speak Spanish so, so I feel more comfortable with Spanish. But, um, English is important so'.

[SD M2]

(21) Cede the Floor

La familia de mi papá es más alejada, so.

'My dad's family is more distant, so'.

[SD L2]

All four functions of *so* are attested in the San Diego second-generation speakers (SD2), but only two of them in the first generation (SD1). In Table 6.25 I present the number of attestations of *so* per function for each of the two groups. The most popular function for both groups is to express cause and/or consequence. *So* is used very infrequently to change topics in both groups. *So* is used only by the second-generation (SD2) speakers to begin a discourse segment and to cede the floor.

Table 6.25 Functions of *so* per Population. Number indicates number of attestations.

	SD1	SD2	Total
Cause consequence	9	14	23
Change topics	2	3	5
Begin a discourse segment	0	4	4
Cede the floor	0	5	5
Total	11	26	37

Lipski (2005) notes that cases of *so* insertion into Spanish speech are more common among speakers who are English dominant, as it is the case for the San Diego second-generation (SD2) speakers. Moreover, he mentions that immigrants who become English dominant as a result of living, working, or studying in the United States also exhibit insertion of *so* (Lipski 2005). This is not the case for the San Diego first-

generation (SD1) speakers in this study, however, since none of them are English dominant. The presence of *so* in their repertoire of discourse markers cannot be ascribed to English dominance but rather to frequent and intense contact with a variety of Spanish that already uses *so* as a discourse marker.

Both San Diego groups use *so* to express cause/consequence and to change topics, but only the second generation (SD2) uses it to begin a discourse segment and to cede the floor. San Diego first-generation speakers (SD1) only adopt a subset of the functions of *so* that second-generation speakers (SD2) use. Not only is the use of *so* reduced among SD1 speakers when compared to SD2 speakers, but also reduced among the older generation is the number of functions. Second-generation speakers (SD2) may represent the conduit through which *so* is transferred from English to Spanish. English-dominant second-generation speakers (SD2) first transfer *so* from English with four functions, and in turn the first-generation Spanish-dominant speakers (SD1) adopt *so* with two of these functions and in smaller numbers through interaction with second- and first-generation speakers who have already adopted it.

6.7.3 Language Experience and the Use of *so* in San Diego

In this section I explore those aspects of the language experience (exposure and interaction) that may be related to the use of *so* among speakers in San Diego. I focus on the sociodemographic data as well as the results from the audio perception study for both populations in San Diego.

I start with the SD2 speakers since they are English dominant and they represent the population with the most widespread use of *so*, relatively speaking, and the majority of their interactions are in English and Spanish. For the SD2 group, the three factors that emerge as related to use of *so* are gender, total number of fillers, and results from the audio perception study. In section 6.7.2 I point out that 5 out of 11 heritage speakers use *so*, which corresponds to 45.4% of the SD2 population. Of these 5 speakers, 4 are male. Therefore, gender appears to be the only relevant sociodemographic factor for the heritage speakers with 80% of the speakers who use *so* being male.

The total number of fillers used by speakers may also be related to use of *so* among heritage speakers. In Table 6.26 I provide the total number of fillers generated by each speaker in SD2, in ascending order from lowest to highest. The total ranges from 31 to 88 total fillers and the median for this is 51 fillers.

Table 6.26 Total Number of Fillers per Speaker (SD2)

Speaker	Uses *so*?	Total number of fillers
SD-I2	no	31
SD-F2	yes	34
SD-C2	yes	36
SD-M2	yes	41
SD-L2	yes	45
SD-K2	no	51
SD-B2	yes	57
SD-A2	no	64
SD-D2	no	68
SD-J2	no	76
SD-H2	no	88

Of the 5 speakers who use *so*, 4 (80%) have a total number of fillers that falls below the median of 51 for the SD2 group. This may indicate that use of *so* in this population is related to a lower overall use of fillers. However, it is important to bear in mind that the use of *so* is already very low frequency for all San Diego populations, therefore the fact that *so* may be related to a lower overall use of fillers in no way suggest that *so* is replacing other fillers.

In the audio perception study, 4 of the 5 (80%) SD2 speakers who use *so* are correctly identified as *americano* (heritage) by at least 5 of the 8 Chihuahua listeners. These results are summarized in Table 6.27, with the column that contains the number of Chihuahua listeners who identify SD2 speakers as *americano* is sorted in descending order.

Table 6.27 Audio Perception Study. SD2 Speakers as *americano*

Speaker	Uses *so*?	Number of Chihuahua listeners (out of 8) who identify SD2 speaker as *americano*
SD-F2	yes	7
SD-L2	yes	7
SD-B2	yes	6
SD-C2	yes	5
SD-D2	no	4
SD-I2	no	4
SD-K2	no	2
SD-A2	no	2
SD-M2	yes	1
SD-H2	no	0
SD-J2	no	0

The four SD2 speakers with the greater number of Chihuahua listeners that identify them as *americano* are users of *so*. However, no fillers in English were present in any of the audio segments that Chihuahua listeners heard. This may signal that SD2 speakers who use *so* may be activating other perceptible features (e.g. intonational) that make them distinguishable as dominant speakers of English. This is particularly true if we accept Lipski's (2005) proposal that the filler *so* in the Spanish of U.S. bilinguals represents an instance of momentary code-switching. However, in the San Diego-Tijuana corpus, *so* is very low-frequency -- as is code-switching -- therefore Lipski's proposal cannot be truly corroborated by the findings of the study.

In sum, heritage speakers in the present corpus who are male and have an overall low use of fillers in their group may exhibit a greater tendency to use *so* in Spanish. Use of *so* for this group is also related to being perceived as *americano* by monolingual Chihauhua listeners even when not using this filler. This suggest that users of *so* may be activating other features in their Spanish that make them sound somewhat different from their SD2 peers who do not use *so*.

The salient factors from the language experience of the San Diego first-generation speakers that are related to use of *so* are age and number of years in the border area. The use of *so* is even less frequent among SD1 than SD2 speakers, so establishing a relationship between the use of this filler and any sociodemographic factors is difficult since only 3 (out of 11) speakers in SD1 use *so*.

Regarding age, the ages of the SD1 speakers range from 41 to 60. The median age for the group is 50, and four of the eleven SD1 speakers are at the median. The three

SD1 speakers who use *so* are below the median age of 50. This may mean that younger parents are more likely to adopt the filler *so*. However, one of the speakers (SD-I1) at 49 years old is only one year away from the median age of 50. As such, a generalization cannot be made about age. Nonetheless, the two youngest parents in the corpus are users of *so*. The ages for the SD1 group are presented in Table 6.28 in ascending order along with other sociodemographic and language experience data that I will discuss below.

Table 6.28 Age of the SD1 Speakers

Speaker	Uses *so*?	Age	Gender	SD1 Child uses *so*?	Gender of SD2 child
SD-C1	Yes	41	F	Yes	M
SD-B1	Yes	46	F	Yes	M
SD-K1	No	47	F	No	F
SD-I1	Yes	49	M	No	M
SD-F1	No	50	F	Yes	M
SD-L1	No	50	M	Yes	F
SD-E1	No	50	F	No	F
SD-J1	No	50	M	No	F
SD-H1	No	52	M	No	M
SD-D1	No	52	F	No	F
SD-M1	No	60	M	Yes	M

The two youngest parents in SD1 are users of *so*, female, and both have male children who use *so*. Only one other female parent in the corpus (SD-F1) has a male child who is a user of *so*. However, SD-F1 has the highest-educational level in Mexico of any other SD1 parent (a law degree) and is also a business owner in the United States. Since her socioeconomic and educational background are higher than the majority of SD1

speakers in the corpus, she may make a conscious effort not to codeswitch or insert English fillers into her Spanish.

The other sociodemographic factor that may be related to use of *so* among SD1 speakers is the number of years that parents have spent in the United States. For all parents in San Diego, this number ranges from 18 to 30 years. This information is provided in ascending order in Table 6.29.

Table 6.29 Number of Years in the Border for SD1 Speakers

Speaker	Uses *so*?	Number of years in the United States
SD-C1	Yes	18
SD-B1	Yes	18
SD-I1	Yes	20
SD-E1	No	20
SD-H1	No	21
SD-F1	No	22
SD-L1	No	24
SD-K1	No	25
SD-M1	No	25
SD-J1	No	28
SD-D1	No	30

The three parents who use *so* are also among the most recent arrivals to the United States, relatively speaking. Because one of the requirements for participation in the study was that parents have an adult child born in the United States, no data is available for any parents who have been in the United States less than 18 years.

For the sake of completion and presenting the results of the audio perception study for both San Diego groups, below I provide the number of Chihuahua listeners

who identify SD1 speakers as U.S. speakers of Spanish, either *inmigrante* or *americano*. This data is presented in Table 6.30 in descending order.

Table 6.30 Audio Perception Study. SD1 Speakers as *U.S. speaker*

Speaker	Uses *so*?	Number of Chihuahua listeners (out of 8) who identify SD1 speaker as *U.S. Speaker* (either *inmigrante* or *americano*)
SD-B1	**Yes**	**8**
SD-J1	No	8
SD-C1	**Yes**	**8**
SD-M1	No	8
SD-I1	**Yes**	**7**
SD-E1	No	6
SD-D1	No	6
SD-K1	No	5
SD-F1	No	5
SD-L1	No	5
SD-H1	No	3

The three SD1 speakers who use *so* are perceived by the majority of Chihuahua listeners (7 and 8 out of 8) as being U.S. speakers of Spanish, either *inmigrante* or *americano*. Unlike the results of the audio perception study for the SD2 group, which presented a relationship between the use of *so* and being perceived as *americano*, the results for SD1 do not show such a clear relationship since non-users of *so* in the SD1 group are also perceived as U.S. speakers of Spanish by a majority of Chihuahua listeners.

Since the number of SD1 speakers who use *so* is very small, and the frequency of use is very low, the results that I provide in this section only establish a descriptive relationship between the language experience data & the audio perception study and the use of *so*. A larger speaker sample may yield more uses of *so* in more speakers that a

statistical analysis may be carried out. As it stands in the current study, it appears that SD1 parents acquire *so* from interaction with their children and at work (the only other significant source of interaction in English and Spanish for SD1). What may facilitate this process is if the parent is female and the child is male. Age and number of years in the U.S. may also play a role in that being a younger parent who has been in the U.S. 18-20 years is related to use of *so* in the corpus.

6.8. The Status of Fillers in the San Diego-Tijuana Border Area

In section 6.4 I outlined four research questions to guide this study. I now revisit the four research questions to summarize the results.

> 1. Is there a core set of fillers in the San Diego-Tijuana border area that is common to all speakers, similar to what Valdés and Geoffrion-Vinci (1998) find? Considering that speakers in their two populations used different strategies to fill pauses beyond the core fillers, are there fillers beyond the core group in San Diego-Tijuana that emerge as significant for different populations?
>
> Yes, there is a core set of fillers in the border area. It is described in (22).
>
> (22) Core fillers in the San Diego-Tijuana border area
> *como que*, *digo*, *eh*, *okay*, and *pues*

Beyond the core set of fillers, the following fillers in (23) emerge as significant for the different populations.

(23) Fillers Statistically Significant for Different Populations
Tijuana: *este*, *¿no?*, *o sea*
San Diego: *ah*, *so*
Non-heritage: *bueno*, *entonces*
Heritage: English fillers such as *like* and *you know*
Immigrant: *¿verdad?*

Whereas Valdés and Geoffrion-Vinci (1998) focus on an academic register for their study, the present study focuses on an informal register where the interactional occasion is conversational and largely autobiographical.

2. Taking into account core fillers for all speakers in addition to the population-specific fillers, how many different sets of fillers are there in the San Diego-Tijuana border area?

There are three sets of fillers in the border area: Tijuana, Immigrant, and Heritage. These are outlined in Table 6.31 below and each consists of the core fillers plus the fillers that are significant for each population.

Table 6.31 Three Sets of Fillers in the Border Area

Set	Types of fillers			
Tijuana	Core fillers	Tijuana fillers	Non-heritage fillers	
Immigrant	Core fillers	San Diego fillers	Non-heritage fillers	Immigrant fillers
Heritage	Core fillers	San Diego fillers	Other English fillers	

3. Since most studies cite the presence of English fillers in the Spanish of U.S.-born Hispanics, what are the English fillers that are present in the border area? Are any English fillers used significantly/consistently by any group(s)?

The two English fillers in the border area are *okay* and *so*. The filler *okay* is present on both sides of the border for both generations and it emerges as one of the core fillers for the area. Its frequency for all groups is low with an overall frequency median of 0. That *okay* surfaces in the Tijuana groups – particularly the first generation (TJ1) whose contact and knowledge of English is for the most part very limited – indicates that *okay* has been lexicalized as a Spanish filler at least in the border area.

The filler *so* is only attested in San Diego speakers. I provide a full summary of the status of this filler in the answer to the next research question. In addition to *so*, a number of English fillers such as *you know* and *like* are part of the repertoire of fillers for the Heritage speakers (SD2) but they are very low in frequency and outside of the scope of this study.

4. What is the status of the English filler *so* in the border area? Is it used by both groups, heritage speakers and adult immigrant speakers to the United States as Silva-Corvalán (1994) finds in Los Angeles Spanish? If so, is it used differently and with different functions across generations in San Diego, as Lipski (2005) and Aaron (2004) find in their studies? Following Torres & Potowski (2008), does an increased use of *so* result in a reduced use of *entonces*? Or is the status of *so* in the border area similar to what Aaron (2004) found for New Mexico where *so* and *entonces* are in free variation and equally stable?

As Silva-Corvalán (1994) finds for Spanish in Los Angeles, *so* is used by heritage speakers (SD2) and adult immigrant speakers (SD1) in San Diego. For both San Diego populations it is low frequency (median = 0) but it is used by more heritage speakers (n=5) than immigrant speakers (n=3). It is also used with more functions by heritage speakers than immigrant speakers. Whereas SD2 speakers use it with four different functions, immigrant speakers use it only with two. This corroborates results from Lipski (2005) and Aaron (2004) in that *so* is used with multiple functions among heritage speakers. However, their studies do not take into account immigrant speakers. Regarding the use of *so* and *entonces,* the filler *entonces* emerges as statistically significant for the non-heritage group. Heritage speakers show a reduced use of *entonces* (median = 2) compared to non-heritage speakers (median = 6). However, because *so* is low frequency for both San Diego groups (median = 0 for both), there is not enough data to determine whether an increased use of *so* among heritage speakers correlates

with a reduced use of *entonces* for the same group. For speakers in San Diego, there is a reduced use of *entonces* for heritage speakers while *so* remains low-frequency for both SD1 and SD2 populations.

6.9 Conclusion

One of the features that differentiates the three sets of fillers in the San Diego-Tijuana border area is the presence of English fillers, all of which are low frequency. While *okay* is part of the core set of fillers for all populations, it is the presence of *so* in San Diego that may best exemplify language change. *So* appears to first be introduced into the set of fillers in the border area by the heritage speakers (SD2), who are English dominant. Once introduced, *so* may become part of the contact variety of Spanish in San Diego with four different functions. Heritage speakers can then complement their set of fillers not only with *so* in low frequency but also with other English fillers in low frequency. Because their Spanish fillers are also low in frequency, the heritage set of fillers is characterized by the low frequency of the majority of its fillers. It is worth pointing out that heritage speakers may reduce fillers in favor of silent pauses and repetitions, but these were not analyzed as part of the present study and remain an avenue for future study. Immigrant speakers (SD1) may incorporate *so* into their f set of fillers through exposure and interaction with the contact variety of Spanish spoken by the heritage speakers (SD2) and more established immigrants (SD1). While it may be possible that some immigrant speakers may transfer *so* directly from English, this is unlikely for the San Diego group in the present study since their exposure to English is

very variable and none of them are English dominant. Immigrant speakers complement their set of fillers with *so* but they do so in very low frequency without reducing the frequency of their Spanish fillers. Moreover, the filler *so* appears to be transferred from the SD2 group to the SD1 speakers with a reduced number of functions.

In a language contact situation in San Diego, the driving force behind language change in the sets of fillers is exposure to and interaction in English. For the English-dominant group (SD2), the use of *so* in Spanish is a transfer effect directly from English. The integration of *so* into the set of the immigrant group (SD1) is due to exposure and interaction with a variety of Spanish that has transferred this filler from English, and possibly as a transfer effect directly from English[3]. That the heritage group (SD2) integrates *so* with a complete set of functions (four) while the immigrant group (SD1) integrate it with a reduced one (two functions) indicates that the filler is not transferred uniformly to all San Diego populations but rather that it originates with the English-dominant group as part of the heritage set of fillers and it is later adopted by the immigrant set with a subset of functions.

[3] The extent to which *so* represents a transfer effect directly from English for the SD1 generation is difficult to ascertain in the present study with the current data given that the only English-only interaction for some SD1 speakers is at their place of employment. While the rest of their interactions may be in Spanish and some English, not all SD1 speakers were specific or forthcoming about their use of English in the interviews.

Chapter 7 Conclusion

In this dissertation I investigated the effects of language contact on the use of subjunctive and fillers among speakers of Spanish in the Tijuana-San Diego border area. I set aside questions of language attrition and incomplete acquisition, and instead focused on formal schooling in Spanish and language dominance through exposure and interactions as the two factors of language contact that are possible causes for the emergence of a heritage lect in San Diego. Without longitudinal data, it cannot be determined whether a language feature has been acquired in childhood and then lost in adulthood. Benmamoun, Montrul, & Polinsky (2010) point out that due to individual speaker variation in heritage communities, language attrition may be very difficult to prove. Since the data for the present study is not longitudinal, I did not pursue the possibility of language attrition being a factor of language change in the border area. Work by Otheguy (2016) and Otheguy & Zentella (2012) has called into question the notion of incomplete acquisition when referring to the lect spoken by U.S.-born second-generation speakers of Spanish. Instead, Otheguy and Zentella (2012) propose that the bilingual lect has evolved differently from a reference lect but it is by no means an incomplete one.

Taking as a point of departure Otheguy & Zentella's (2012) notion that a bilingual lect is not incomplete but rather that it is different, I set out to determine how the four populations under study on both sides of the border differ in terms of proficiency, exposure and interactions in English and Spanish, and formal schooling in Spanish. In regards to proficiency, all four groups perform equally in speech rate and

inventory of nouns. In terms of their inventories of verbs, the SD2 group shows a statistically significant difference from the rest of the speakers (non-heritage) in that they (heritage speakers) operate with a somewhat reduced repertoire of verbs when compared with the non-heritage speakers. In an audio perception study, monolingual listeners in Chihuahua, Mexico, perceived three different categories of speakers in the border: Tijuana speakers, immigrant speakers, and San Diego/heritage speakers. In spite of these perceptual differences and only differing from the non-heritage group in their unique verb use, heritage speakers in San Diego represent a very highly-functional and highly-proficient group of speakers of Spanish.

Regarding exposure and interactions in Spanish and English, the Tijuana speakers are completely Spanish dominant though some of them report some proficiency in English as a result of schooling as well as exposure to U.S. Spanish through family and friends in the San Diego area. The SD1 speakers are also Spanish dominant, but they report exposure to English through some media. Their interactions are for the most part in Spanish, but as a group they exhibit variation in their interactions in English: some of them use English at work and in some social situations whereas other report not using it at all. It is important to emphasize that their most significant interactions, with family, are in Spanish. Another aspect of their language experience where SD1 speakers exhibit great variation is in their reported schooling in English. A number of SD1 speakers have taken night classes or adult-school courses to learn English, or have done so through at-home self-guided materials.

SD2 speakers, on the other hand, are English dominant as the majority of their interactions with siblings, with friends, at school, and generally outside the home, are primarily in English. They report some exposure to Spanish through some media. Their interactions in Spanish are primarily with family members at home and with older family members outside the home. SD2 speakers are native speakers of Spanish since all report Spanish as their first language. In spite of being English dominant and demonstrating great variation in the language of their interactions, SD2 speakers have grown up in an environment that generates a competent grammar in Spanish.

Concerning access to formal schooling in Spanish, the first-generation speakers on both sides of the border (TJ1 and SD1) exhibit a low to medium level of schooling in Spanish. The second-generation speakers in Tijuana exhibit the highest. Heritage speakers in San Diego report little to no access to formal schooling in Spanish. What is remarkable about this heritage population is that their lect is shaped almost exclusively by virtue of being native speakers and growing up in a home environment where Spanish is spoken. As I remarked above, the SD2 speakers are highly proficient in Spanish in spite of their lack of formal schooling in the language.

Two language features were analyzed to determine the effects of language contact on Spanish in San Diego: the use of subjunctive and the use of fillers. Regarding the subjunctive, there was no statistically significant difference in overall production of subjunctive for the four populations. However, as subjunctive is sometimes used in conditional contexts, taking into account only subjunctive contexts reveals that heritage speakers performed in a manner that is statistically significantly different from non-

heritage speakers. In other words, heritage speakers reduced their use of subjunctive only in subjunctive contexts. In overall use of subjunctive, they performed no different from non-heritage speakers.

In obligatory contexts (S-ONLY), heritage speakers produced more verbs in the present subjunctive form that non-heritage speakers did. However, the difference was marginally statistically significant. In alternating contexts (S-ALT), it was the Tijuana second-generation population that performed in a manner that was statistically significantly different from the rest of the speakers in that they (TJ2) generated more S-ALT contexts than the other groups. In COND-EXP contexts – non-subjunctive contexts where the conditional is sometimes used – the SD2 group generated the most forms whereas the TJ2 population generated the least of the four groups.

Regarding subjunctive use in the border, the two groups that showed the most significant differences were the TJ1 and SD2 groups. Recall that the TJ1 speakers are the ones with the highest level of formal education in Spanish. They were also the ones who demonstrated a higher use of S-ALT contexts and the lowest use of COND-EXP contexts. This may suggest that formal schooling in Spanish exposes speakers to more forms and contexts where the subjunctive and the indicative alternate (S-ALT contexts). Given that there is a pragmatic difference between subjunctive and indicative use in S-ALT contexts, it may be the case that a higher education level in Spanish provides speakers with more opportunities where these contexts can be used. Another effect that formal schooling in Spanish has on the subjunctive is in COND-EXP contexts. TJ2 speakers generated the least of these context and SD2 the most. In the case of COND-EXP

contexts, the determining factors are not only education in Spanish but also language dominance since SD2 is the only group that is English-dominant.

Not only did the SD2 group produce a greater proportion of COND-EXP contexts, they also generated a greater proportion of S-ONLY contexts. Because SD2 are the only group that is English dominant and with no formal schooling in Spanish, this may suggest that the extent of their interactions in Spanish in San Diego affords them the opportunity to use more s-only contexts. However, since the difference is marginally statistically significant I make this claim with some reservation. Use of subjunctive is COND-EXP contexts is a feature that is present in all four groups in the border, but it is significantly more prominent in SD2 speakers. That it is least prominent in the group with the most formal education in Spanish (TJ2) and most prominent in the group with the least formal schooling in Spanish (SD2) suggests two things: (1) that on the one hand, COND-EXP contexts may be disfavored in certain registers (professional or academic) where a more standardized variety of Spanish is preferred; (2) on the other hand, English dominance in a U.S. variety of Spanish may favor this form. The prevalence of COND-EXP forms among SD2 speakers of Spanish may be evidence that changes already present in a non-contact variety of a language become accelerated in a language contact situation (Silva-Corvalán 1994). The data from this present study underscores the lack of formal schooling in Spanish and English dominance as important components of a language contact situation.

With regards to the filler system, there is a core set of fillers that is used by all speakers in the border area. In addition, some subsets emerge as statistically significant:

Tijuana, San Diego, non-heritage, heritage, immigrant, and English. A set of core fillers plus certain subsets result in three systems of fillers in the border area:

1. Tijuana (core fillers, Tijuana fillers, and non-heritage fillers)
2. immigrant (core fillers, San Diego fillers, non-heritage fillers, immigrant fillers)
3. heritage (core fillers, San Diego fillers, heritage fillers, and English fillers)

Work by Matras (1998) and Fuller (2001) has pointed out that fillers are prone to borrowing in situations of language contact. In the border area, the English fillers *so*, *like*, and *you know* are borrowed into the Spanish of the SD2 group. Considering that these speakers are English dominant, the borrowing of these three English fillers by SD2 speakers represents a mixed system (Fuller 2001) that is typical of a variety of Spanish in the United States where the pragmatically dominant language (Matras 2998) is English.

The presence of the English filler *so* in the SD1 population cannot be attributed to English dominance given that the SD1 group is Spanish dominant. What this indicates is that the pragmatically dominant language (Matras 1998) for the SD1 speakers is neither English nor the same variety of Spanish spoken in Tijuana. Rather, their pragmatically dominant language is a lect of U.S. Spanish that has already undergone changes. Because *so* is very infrequent in the speech of the SD1 group, and it is the only English filler borrowed by SD1, its presence in Spanish does not indicate that English fillers may be replacing Spanish fillers. Rather, SD1 immigrant speakers possess an

emerging mixed system of fillers from a pragmatically dominant San Diego lect. In regards to exposure and interactions, the SD2 speakers first borrow the English fillers into their mixed system where English is the pragmatically dominant language. The SD1 group then borrows only *so* from the SD2 system into an SD1 mixed system where the programmatically dominant language is a San Diego lect of Spanish. It is worth nothing here that Matras (1998) uses the term 'pragmatically dominant language' in a community so as to separate literacy from language use.

San Diego second-generation speakers of Spanish (SD2), in spite of being a highly proficient group in Spanish, demonstrate an attenuated use of subjunctive only in subjunctive contexts. The likely causes for this difference in control of the subjunctive mood when compared to non-heritage populations are English dominance in interactions as well as lack of formal schooling in Spanish. English dominance may also account for the borrowing of English fillers into their mixed system of fillers. Their parents, the SD1 group, through prolonged interaction and exposure to the pragmatically dominant language in their community, the San Diego Spanish lect, develop their unique mixed system.

The present study is unique among studies of heritage and immigrant speakers of Spanish in that it incorporates Tijuana speakers not as native controls but as points of reference for a grammar that is comparable to what immigrant speakers possessed before immigrating to San Diego. The sociodemographic background of the Tijuana parents is similar to the San Diego parents', and as such speakers in San Diego are not compared to a group of speakers who are radically different from them. In other words,

San Diego heritage speakers, who are native speakers of Spanish , have grown up in an environment where they have acquired a competent grammar that is comparable to their parents' and to Tijuana speakers'.

I conclude this dissertation by presenting some possibilities for further investigation as a result of the findings from this project:

1. A study of silent pauses and repetitions among heritage speakers to determine if their overall reduction of fillers is due to an increase in silent pauses, false starts, and repetitions. An investigation of this phenomena for all groups in the border area can provide a more complete description of the subjunctive system.
2. A more in-depth investigation on the effects of formal schooling among Spanish monolinguals, particularly monolingual speakers with little to no formal schooling in Spanish. This would allow for a comparison between two populations with no formal schooling in Spanish but a completely different pattern of interactions. The use of subjunctive and fillers in a monolingual no-literacy setting can reveal which features of heritage Spanish are different due to no formal schooling in the language.
3. An audio perception study that includes not only Mexican monolinguals outside of the border area but also Tijuana speakers, immigrant and heritage speakers in San Diego. Furthermore, this study would present listeners with fewer audio clips as well as with a different scale to categorize speakers in more than three categories or along a continuum.

The Tijuana-San Diego border area presents a unique sociolinguistic setting that is ideal for the study of language contact. In this dissertation I have described two features of a heritage language system and proposed that exposure and interactions in Spanish and English, as well as lack of formal schooling in Spanish, are two significant forces behind language contact in the border area.

Appendix 1 Interview Questions

Below is list of the questions asked to all participants in this study. Because different interviewers were involved over the course of two years, not all of the questions were asked of all the subjects in the exact same way. I provide an English translation for each of the questions

1. Sociodemographics

Díganos sus datos personales: su nombre, edad, de dónde es, a qué se dedica, dónde vive, y cuántos años tiene viviendo aquí.

> 'Tell us some of your personal data: name, age, where you are from, what you do for a living, where you live, and how long you've been living here'.

¿Quiénes son los miembros de su familia? Díganos sus nombres, edades, y cuál es el parentesco.

> 'Who are your family members? Tell us their names, age, and their relationship to you.'

¿A qué se dedica usted y cuáles trabajos ha tenido, y más o menos por cuánto tiempo?

> 'What do you do for a living and what jobs have you had and more or less for how long have you had them?'

¿En qué lugares ha vivido y más o menos por cuánto tiempo en cada uno?

> 'Where else have you lived and more or less for how long have you lived in each place?'

2. Language Use

(SD ONLY) *Con su familia, ¿qué tanto habla español e inglés, y con quién habla cada idioma?*

> 'With your family, how much Spanish and English do you speak? Who do you speak Spanish or English with?'

(SD ONLY)*¿En qué situaciones usa más el español y en qué situaciones usa más el inglés: con su familia, en el trabajo, con sus amigos, en situaciones sociales (iglesia, fiesta), al leer, escribir, ver televisión, escuchar la radio? ¿Qué tipo de cosas ve, lee, escribe, escucha en español/inglés?*

> 'In what situations do you use Spanish more? In what situations do you use English more? What about with family, work, friends, in social situations (church, parties), reading, writing, television, radio? What kinds of things do you watch on TV, read, write, listen to in Spanish/English?'

(SD ONLY) *¿Cómo determina usted si va a hablar en español o en inglés con alguna persona o en alguna situación que usted no conoce, por ejemplo al ir a una tienda donde nunca ha ido, o al encontrar una persona que nunca ha conocido?*

> 'How do you decide if you're going to speak in Spanish or in English with someone you don't know, or if you're in an unfamiliar situation? For example, if you go to a store you've never been to before, or if you meet someone for the first time'.

(TJ ONLY) *¿Ha estudiado usted algún otro idioma? ¿Por cuánto tiempo? ¿Qué porcentaje tiene y cuándo lo usa?*

> 'Have you studied another language? For how long? How would you rate your proficiency and when do you use this other language?'

(SD First-generation only) *¿Cuál es su nivel de inglés? ¿Cuándo lo ha estudiado y cómo, y por cuánto tiempo?*

> 'What's your level of English? When, where, and how have you studied it, and for how long?'

(SD Second-generation only) *¿Cuál es tu nivel de español? ¿Cuándo lo has estudiado y cómo, y por cuánto tiempo?*

> 'What's your level of Spanish? When, where, and how have you studied it, and for how long?'

3. Personal Experiences and Opinions

Descríbanos el lugar de dónde usted es.

> 'Describe the place you are from'.

¿Cómo describiría usted San Diego/Tijuana a una persona que no conoce este lugar?

> 'How would you describe San Diego/Tijuana to someone who has never been here?'

¿Cuál considera usted que es uno de los momentos más felices de su vida? ¿Qué pasó, cuándo fue, y por qué lo considera tan importante?

> 'What is one of the happiest moments of your life? What happened? When did it happen, and why do you consider it important?'

¿Cuál considera usted que fue un momento triste o vergonzoso en su vida? ¿Qué pasó, cuándo fue?

> 'What is a sad or embarrassing moment from your life? What happened and when?'

(SD ONLY) *¿Cómo cree usted que sería su vida diferente si viviera del otro lado de la frontera y jamás hubiera vivido en Estados Unidos?*

'How do you think your life would be different if you lived on the other side of the border and you had never come to the United States? '

(TJ ONLY) *¿Cómo cree usted que sería su vida diferente si viviera del otro lado de la frontera en Estados Unidos?*

'How do you think your life would be different if you lived on the other side of the border, in the United States?'

¿Qué les desea usted a las generaciones de los jóvenes, para el futuro? ¿Cómo le gustaría que las cosas fueran a futuro?

'What do you wish for future generations? How would you like for things to be in the future (for young people)?'

4. Language Attitudes

(TJ ONLY) *¿Qué opina usted del español que se habla aquí en la frontera? ¿Y qué opina usted del español que hablan las personas del otro lado que son nacidas allá?*

'What do you think about the Spanish that is spoken here in the border? And what do you think about the Spanish that is spoken in the United States by people who were born there?'

(SD ONLY) *¿Qué opina usted del español que se habla aquí en San Diego por las personas que son nacidas aquí? ¿Nota usted algunas diferencias entre el español de los jóvenes de aquí, que son nacidos aquí, y el español de los padres de estos jóvenes?*

'What do you think about the Spanish that is spoken here in San Diego by people who were born here? Do you notice any differences between the way young people speak, those who were born here, and their parents?'

(SD First generation only) *Y como usted es de México, ¿qué opina del español que se habla aquí comparado al español que se habla allá?*

'Since you are from Mexico, what do you think about the Spanish that is spoken here compared to the one that is spoken over there?'

(SD Second generation only) *¿Qué opinas del español que se habla aquí, del español que tú y los jóvenes de tu edad hablan?*

'What do you think about the Spanish that is spoken here, the kind that you and your friends your age speak?'

Appendix 2 Demographic Information

Below is the table for the demographic information for the 44 participants selected for the study. Gaps in the ABC labeling sequence are the result of either incomplete interview sets, only one of the family members being available for the interview, or an interview set being discarded due to insufficient length.

Table 8.1 Demographic Information

Label	Gender	Age	Place of Origin	Current Residence	Years in Border Area
SD-A2	F	18	Los Angeles	National City	18
SD-B1	F	46	Chihuahua	National City	18
SD-B2	M	18	National City	National City	18
SD-C1	F	41	Tijuana	National City	18
SD-C2	M	21	Los Angeles	National City	21
SD-D1	F	52	Queretaro	National City	28
SD-D2	F	18	San Diego	National City	18
SD-E1	F	50	Oaxaca	National City	20
SD-F1	F	50	Sinaloa	National City	22
SD-F2	M	18	Chula vista	National City	18
SD-H1	M	52	Mexicali	National City	21
SD-H2	M	18	National City	National City	18
SD-I1	M	49	Jalisco	National City	20
SD-I2	M	18	National City	National City	18
SD-J1	M	50	Jalisco	National City	28
SD-J2	F	25	National City	National City	25
SD-K1	F	47	Sinaloa	National City	25
SD-K2	F	20	San Diego	National City	20
SD-L1	M	50	Jalisco	Lemon Grove	24
SD-L2	F	20	San Diego	Lemon Grove	20
SD-M1	M	60	Jalisco	National City	25
SD-M2	M	20	National City	National City	20
TJ-B2	F	22	Tijuana	Tijuana	22
TJ-F1	F	54	Guanajuato	Tijuana	40
TJ-L1	M	69	Colima	Tijuana	48
TJ-L2	M	32	Tijuana	Tijuana	32
TJ-M1	M	57	Sonora	Tijuana	22
TJ-M2	F	23	Sonora	Tijuana	20
TJ-O1	M	43	Tijuana	Tijuana	43
TJ-O2	M	39	Tijuana	Tijuana	18
TJ-P1	F	39	Tijuana	Tijuana	39
TJ-P2	F	19	Tijuana	Tijuana	19

Table 8.1 Demographic Information (continued)

Label	Gender	Age	Place of Origin	Current Residence	Years in Border Area
TJ-Q1	M	37	Mexico DF	Tijuana	15
TJ-Q2	F	18	Tijuana	Tijuana	18
TJ-R1	F	46	Morelos	Tijuana	16
TJ-R2	M	18	Morelos	Tijuana	18
TJ-S1	F	44	Sonora	Tijuana	23
TJ-S2	F	21	Tijuana	Tijuana	21
TJ-T1	F	36	Coahuila	Tijuana	14
TJ-T2	M	38	Coahuila	Tijuana	14
TJ-U1	M	40	Coahuila	Tijuana	15
TJ-U2	F	18	Coahuila	Tijuana	15
TJ-V1	M	54	Jalisco	Tijuana	51
TJ-V2	M	18	Tijuana	Tijuana	18

Appendix 3 Codes

Once all interviews were transcribed, they were broken down into lines so that each line of text would contain only one verb (finite or non-finite). Each verb was then coded in Excel according to the criteria outlined in Table 8.2.

Table 8.2 Criteria for Coding Verb Forms

Category	Description and Codes
FINITE Finite or non finite form	FIN: finite verb NONFIN: non-finite verb INCOMP: incomplete form of the verb * if a verb is coded as IMCOMP, no further category is coded for it
MOOD Mood	INDIC: indicative SUBJ: subjunctive IMP: imperative COND: conditional NEG-IMP: negative imperative * if verb is non-finite, MOOD is left blank
FORM Verb form	PRES: present PRET: preterit IMPERF: imperfect FUT: future COND: conditional INFIN : infinitive PRESP : present participle or gerund PP : past participle *if MOOD is IMP or NEG-IMP, this cell for FORM is left blank
PERSNUM Person/number	1SG: first-person (*yo*) 2SG-FAM: second-person familiar (*tú*) 2SG-FORM: second-person formal (*usted*) 3SG: third-person singular (*ella*, *él*, impersonal) 1PL: first-person plural (*nosotros*, *nosotras*) 2PL: second-person plural (*ustedes*) 3PL: third-person plural (*ellos*, *ellas*)
REP Repetition	R: repetition NR: no repetition FS: false start HES: hesitation
ERROR Error	ER: error NS : non-standard NE : no error
VERB Verb	Provide the infinitive form of the verb

When coding for person and number (category: PERSNUM) a verb may be ambiguous between the third-person singular (code: 3SG) and the second-person formal *usted* (code: 2SG-FORM) in a present-tense form, such as in (1).

(1) *Si habla español…*
'if she/he/you (formal) speak(s) Spanish…'

In such case, the both possibilities for person and number (PERSNUM: 2SG-FORM/3SG) are coded for the verb *habla*. For a verb in the present subjunctive mood (FINITE: FIN, MOOD: SUBJ, FORM: PRES), such as (2), three possibilities for person and number (PERSNUM: 1SG/2SG-FORM/3SG) are possible.

(2) *para que no hable con nadie*
'so that I/she/he/you(formal) won't speak to anyone'

As with (1), all three possibilities would be coded for the verb *hable*. When coding for repetitions (category: REP), the majority of the verbs have the value NR for non-repetition. When the verb is part of a hesitation form such as *o sea*, its value is HES). Finally, for errors (category: ERROR) the majority of the verbs have the value NE for non-error. The most common non-standard form is *haiga* (FINITE: FIN, MOOD: SUBJ, FORM: PRES, PERSNUM: 1SG/2SG-FORM/3SG, REP: NR, ERROR: NS, VERB: haber) which corresponds to the present subjunctive form of *haber*.

A total of 26,327 verb forms were coded for the entire corpus. For the Tijuana first-generation group (TJ1) there are 9233 total verbs. For the Tijuana second-generation there (TJ2) are 5903 verbs. For the San Diego first-generation (SD1) there are 6340 verbs. And for the San Diego second-generation (SD2) there are 4851 verbs.

Coding for Demographic Information

Each interview was coded for demographic information according to the criteria presented in Table 8.3.

Table 8.3 Criteria for Coding Demographic Information

Category	Description and Codes
SD/TJ City	SD: San Diego TJ: Tijuana
HER Heritage	HER: Heritage NONHER: Non-Heritage * Non-heritage refers to speakers in the SD1, TJ1, and TJ2 groups
GENDER Gender	M: male F: female

Table 8.3 Criteria for Coding Demographic Information (continued)

Category	Description and Codes
GENER Generation	1: first generation 2: second generation
AGE Age at the time of interview	18-60+
YEARSBORD Years in the border area	18-60+
SE-BACK Socioeconomic background	W: working class B: business owner P: professional
OCC Current occupation	S: student W: working class O: office worker B: business owner H: homemaker P: professional
EDU Highest educational level	P: primary, not finished PF: primary, finished M: middle school, or Mexican *secundaria*, not finished MF: middle school, or Mexican *secundaria*, finished H: high school or Mexican *preparatoria*, not finished or in progress HF: high school or Mexican *preparatoria*, finished HG: high school GED C: college, or Mexican *licenciatura*, not finished or in progress CF: college, or Mexican *licenciatura*, finished G: post graduate, or Mexican *maestría*, not finished or in progress GF: post graduate, or Mexican *maestría*, finished

References

Aaron, J. E. (2004). So respetamos un tradición del uno al otro. *Spanish in Context*, *1*(2), 161-179.

Benmamoun, E., Montrul, S., & Polinsky, M. (2010). Prolegomena to heritage linguistics.

Blake, R. J. (1985). From research to the classroom: Notes on the subjunctive. *Hispania*, *68*(1), 166-173.

Bookhamer, K. 2013. *The Variable Grammar of the Spanish Subjunctive in Second generation Bilinguals in New York City. NY*. The City University of New York dissertation.

Borgonovo, C., De Garavito, J. B., & Prévost, P. (2008). Methodological issues in the L2 acquisition of a syntax/semantics phenomenon: How to assess L2 knowledge of mood in Spanish relative clauses. In *Selected Proceedings of the 10th Hispanic Linguistics Symposium*. Ed. Joyce Bruhn de Garavito and Elena Valenzuela. Somerville: Cascadilla (pp. 13-24).

Carreira, M. (2004). Seeking explanatory adequacy: A dual approach to understanding the term heritage language learner. *Heritage Language Journal*, 2(1).

Collentine, J. (1995). The development of complex syntax and mood-selection abilities by intermediate-level learners of Spanish. *Hispania*, 122-135.

--- (2010). The acquisition and teaching of the Spanish subjunctive: An update on current findings. *Hispania*, *93*(1), 39-51.

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, *45*(1), 182-193.

de Leeuw, E., Schmid, M. S., & Mennen, I. (2010). Perception of foreign accent in native speech. *Bilingualism: Language and Cognition*, *13*(1), 33-40.

Erickson, F and Schultz, J. 1981. When is a context? Some issues and methods in the analysis of social competence. In: J. Green and C. Wallat, eds., *Ethnography and language in educational settings*, 147-160. Norwood, NJ: Ablex.

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Reply: Measuring variability in early child language: Don't shoot the messenger. *Child development*, *71*(2), 323-328.

Fuller, J. (2001) The Principle of Pragmatic Detachability in Borrowing: English-original discourse markers in Pennsylvania German. *Linguistics* 39(2): 351-369.

Gallego, M., & Alonso-Marks, E. (2014). Subjunctive use variation among monolingual native speakers of Spanish: A cross-dialect analysis. *Spanish in Context*, *11*(3), 357-380.

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*(03), 433-463.

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

--- (1994). Irrealis and the subjunctive. *Studies in language*, *18*(2), 265-337.

Gudmestad, A. (2008). Acquiring a variable structure: An interlanguage analysis of second-language mood use in Spanish. *Studies in Hispanic and Lusophone Linguistics*, *1*(2), 449-449.

Gumperz, J. J. (1992). 8 Contextualization and understanding. *Rethinking context: Language as an interactive phenomenon*, (11), 229.

Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 3(1), 359– 383.

Lipski, J. (2005). Code-switching or borrowing? No sé so no puedo decir, you know. In *Selected proceedings of the second workshop on Spanish sociolinguistics* (pp. 1-15).

Lynch, A. 1999. *The subjunctive in Miami Cuban Spanish: Bilingualism, contact, and language variability*. MN: The University of Minnesota dissertation.

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal*, *76*(2), 190-208.

Matras, Y. (1998). Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36(2), 281–331.

Menegotto, A. C. (2005). *Morfología verbal del español del Río de la Plata*. Finisterre editores.

Montrul, S. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition* (Vol. 37). John Benjamins Publishing

--- (2009). Knowledge of tense-aspect and mood in Spanish heritage speakers. *International Journal of Bilingualism*, *13*(2), 239-269.

--- (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, 30(3), 3-23.

Montrul, S., & Potowski, K. (2007). Command of gender agreement in school-age Spanish-English bilingual children. *International Journal of Bilingualism*, *11*(3), 301-328.

Ocampo, F. 1990. El subjuntivo en tres generaciones de hablantes bilingües. In John J. Bergen (ed.), *Spanish in the United States: Sociolinguistic issues*. Washington, D.C.: Georgetown University Press. 39-48.

Otheguy, R. (2016). The linguistic competence of second-generation bilinguals: A critique of 'incomplete acquisition.' In *Romance Linguistics 2013. Selected papers from the 43rd Linguistic Symposium on Romance Languages (LSRL), New York, 17-19 April, 2013*. Eds. Tortora, Christina, Marcel den Dikken, Ignacio L. Montoya and Teresa O'Neill [RLLT 9]. John Benjamins Publishing Company (pp. 301-320).

Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity* (Vol. 203). OUP USA.

Polinsky, M. (1997). American Russian: Language loss meets language acquisition. *Formal Approaches to Slavic Linguistics*. Ann Arbor: Michigan. Slavic Publications, 340-370.

--- (2006). Incomplete acquisition: American Russian. *Journal of Slavic Linguistics*, *14*(2), 191-262.

--- (2008). Without aspect. In G. Corbett and M. Noonan, eds. *Case and Grammatical Relations*. 263–282.

--- (2008b). Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization. *Heritage Language Journal*, *6*(1), 40-71.

Polinsky, M., & Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass*, *1*(5), 368-395.

Potowski, K. (2003). Chicago's" Heritage Language Teacher Corps": A model for improving Spanish teacher development. *Hispania*, 302-311.

Potowski, K., & Carreira, M. (2004). Teacher development and national standards for Spanish as a heritage language. *Foreign Language Annals*, *37*(3), 427-437.

Quer, J. (2001). Interpreting mood. *Probus*, *13*(1), 81-111.

Relano Pastor, A. M. (2007). On border identities.'Transfronterizo' students in San Diego. *Diskurs Kindheits-und Jugendforschung*, *2*(3), 263-277.

Schmid, M. S. (2009). On L1 attrition and the linguistic system. *Eurosla Yearbook*, *9*(1), 212-244.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in second language acquisition*, *26*(02), 173-199.

Serrano Montesinos, M. J. (1992). El subjuntivo en Canarias y América.

Silva, G. 2008. Heritage language learning and the Portuguese subjunctive. *Portuguese Language Journal* 3.

Silva-Corvalán, C. 1989. Sociolingüística. Teoría y análisis. Madrid: Alhambra.

--- 1994. Language Contact and Change. Spanish in Los Angeles. Oxford: Clarendon.

--- 1995. The gradual loss of mood distinctions in Los Angeles Spanish. Language Variation and Change 6. 255-272.

--- 2001. Sociolingüística y pragmática del español. Washington, D.C.: Georgetown University Press.

--- 2003. Linguistic consequences of reduced input in bilingual first language acquisition. In *Linguistic theory and language development in Hispanic languages*. Somerville, MA: Cascadilla Press. 375–397.

Tavizón, J. M. (2014). The Spanish Language Proficiency of Sequential Bilingual Children and the Spanish-English Language Proficiency Scale.

Torres, L. 1997. Puerto Rican discourse. A sociolinguistic study of a New York suburb. New Jersey: Lawrence Erlbaum Associates.

-- (2002). Bilingual discourse markers in Puerto Rican Spanish. *Language in Society*, *31*(01), 65-83.

Torres, L., & Potowski, K. (2008). A comparative study of bilingual discourse markers in Chicago Mexican, Puerto Rican, and MexiRican Spanish. *International Journal of Bilingualism*, *12*(4), 263-279.

Valdés, G. (1997). The teaching of Spanish to bilingual Spanish-speaking students: Outstanding issues and unanswered questions. *La enseñanza del español a hispanohablantes: Praxis y teoría*, 8-44.

--- (2000). Bilingualism and language use among Mexican Americans. In S. L. McKay & S. L. Wong, eds. New immigrants in the United States. Cambridge University Press.

Valdés, G., & Geoffrion-Vinci, M. (1998). Chicano Spanish: The problem of the "underdeveloped" code in bilingual repertoires. *The Modern Language Journal*, *82*(4), 473-501.

Vickers, C., & Goble, R. (2011). Well, Now, Okey Dokey: English Discourse Markers in Spanish-Language Medical Consultations. *Canadian modern language review*, *67*(4), 536-567.

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication*, *50*(2), 81-94.

Zentella, A. C. (1997). *Growing up bilingual: Puerto Rican children in New York*. Wiley-Blackwell.

--- (2009). *Multilingual San Diego: Portraits of Language Loss and Revitalization*. University Readers.