HERITAGE SPEAKERS’ ONLINE PROCESSING OF THE SPANISH SUBJUNCTIVE: A COMPREHENSIVE USAGE-BASED STUDY

A Dissertation in

Spanish and Language Science

by

Priscila Lopez-Beltran Forcada

© 2021 Priscila Lopez-Beltran Forcada

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

August 2021
The dissertation of Priscila Lopez-Beltran Forcada was reviewed and approved by the following:

Paola E. Dussias  
Professor of Spanish, Linguistics and Psychology  
Dissertation Advisor

Matthew T. Carlson  
Associate Professor of Spanish and Linguistics

Karen Miller  
Associate Professor of Spanish and Linguistics

Michael T. Putnam  
Associate Professor of German and Linguistics

Damián Vergara Wilson  
Associate Professor of Spanish  
Special Member
ABSTRACT

Studies assessing the grammatical knowledge of speakers of Spanish as a heritage language have largely focused on the Spanish subjunctive mood and have concluded, almost unanimously, that heritage speakers’ knowledge of the Spanish subjunctive is non-native-like and subject to incomplete acquisition. However, there is also evidence that while different, heritage speakers’ linguistic knowledge is by no means deficient. The goal of the present dissertation is to achieve a holistic understanding of the nature of grammars of heritage speakers and to contribute to a theory of processing in heritage language contexts that has greater explanatory adequacy. To this end, this dissertation examines knowledge of the Spanish subjunctive in heritage speakers who live in a long-standing bilingual community in Albuquerque, New Mexico, in comparison to a group of Spanish-dominant bilinguals born in Mexico. The dissertation sets out to provide (1) an evidence-based characterization of heritage speakers by using a sociolinguistic questionnaire which, along with PCA, examined the language-experience related factors that best explain the variability in the processing of the subjunctive mood in this population; and (2) an examination of heritage speakers’ and Spanish-dominant bilinguals’ processing of the Spanish subjunctive during online comprehension and production by means of psycholinguistic experiments that integrated corpus data into their design. Results indicated that, both in comprehension and production, the current group of heritage speakers was sensitive to the lexical and structural conditioning of mood selection, and that the performance of heritage speakers and Spanish-dominant
bilinguals converged on the same results and trends. All participants showed nuanced knowledge of the morphosyntactic factors that modulate the conditioning of mood selection, as suggested by the fact that linguistic factors such as frequency and proficiency also modulated their sensitivity. In addition, based on the PCA conducted, the role of three sociolinguistic variables was examined: use of the heritage language, language entropy, and identification with the heritage language. As predicted, results indicated that sensitivity to the lexical and structural conditioning of mood selection was greater for heritage speakers who: (1) used the heritage language more often on average, (2) used the heritage language in more diverse contexts, and (3) felt more identified with the heritage language. The findings highlight that factors such as the community examined and the ecological validity of the materials used are crucial. In addition, they underscore the importance of triangulating both comprehension and production experimental data, and employing multiple explanatory variables for a more comprehensive approach to complex and highly variable systems such as heritage grammars.
# TABLE OF CONTENTS

**LIST OF TABLES** ........................................................................................................... vii

**LIST OF FIGURES** .......................................................................................................... ix

**ACKNOWLEDGEMENTS** ................................................................................................. xii

**CHAPTER 1: INTRODUCTION** ....................................................................................... 1

- Approaches to the study of heritage grammars ............................................................ 3
  - Traditional approaches: incomplete acquisition and attrition ............................... 5
  - Current approaches: input quality and lexicalist/activation ................................... 8
  - Applying usage-based approached to the study of heritage grammars .................. 11
  - Recapitulation ........................................................................................................... 12

- The linguistic knowledge of heritage speakers ......................................................... 13
  - Findings on phonology ......................................................................................... 13
  - Findings on morphosyntax .................................................................................... 17
  - Task effects ............................................................................................................ 22
  - Recapitulation ........................................................................................................ 24

- The subjunctive mood ................................................................................................. 25
  - Origins of the subjunctive ..................................................................................... 27
  - Traditional characterizations of the subjunctive ................................................... 28
  - Variationist characterizations of the subjunctive .................................................. 31
  - Acquisition of the subjunctive ............................................................................... 35
  - Bilingual grammars and the Spanish subjunctive ................................................... 38
  - Recapitulation ........................................................................................................ 54

- The Spanish of New Mexico and its heritage speaker community ......................... 54
  - Recapitulation ........................................................................................................ 57

**CHAPTER 2: MAXIMIZING THE EXPLANATORY POWER OF QUESTIONNAIRE DATA FOR THE STUDY OF HERITAGE POPULATIONS** ........................................................................................................ 58

- Introduction ............................................................................................................... 58

- Method ......................................................................................................................... 66

- Results ......................................................................................................................... 90

**CHAPTER 3: INVESTIGATING SENSITIVITY TO THE LEXICAL AND STRUCTURAL CONDITIONING OF MOOD SELECTION** ........................................................................................................ 118

- Introduction ............................................................................................................... 118
  - Group 1: Heritage Speakers .................................................................................. 126
LIST OF TABLES

Table 1-1. Variability in definition of obligatory contexts of use across studies.........................................................30

Table 1-2. Effect of non-lexical factors on mood selection.................................................................35

Table 2-1. Education level of participants.................................................................................................67

Table 2-2. Participants’ self-reported characteristics..................................................................................68

Table 2-3. Self-reported proficiency (7-point Likert scale)........................................................................69

Table 2-4. Diversity of language use across social contexts.................................................................69

Table 2-5. Characteristics of picture naming items.....................................................................................72

Table 2-6. Syllable distribution in EIT........................................................................................................73

Table 2-7. Dummy coding of qualitative variables.....................................................................................87

Table 2-8. Eigenvalues of the correlation matrix – Heritage Language Background.........................91

Table 2-9. Correlation of variables with PCs – Heritage Language Background.................................93

Table 2-10. Eigenvalues of the correlation matrix – Contexts of Use....................................................94

Table 2-11. Correlation of variables with PCs – Contexts of Use.........................................................95

Table 2-12. Eigenvalues of the correlation matrix – Productivity of Network........................................96

Table 2-13. Correlation of variables with PCs – Productivity of Network.............................................97

Table 2-14. Eigenvalues of the correlation matrix – Bilingualism of Network....................................................98

Table 2-15. Correlation of variables with PCs – Bilingualism of Network......................................................99

Table 2-16. Eigenvalues of the correlation matrix – Attitude, Motivation & Identity............................100

Table 2-17. Correlation of variables with PCs - Attitude, Motivation & Identity..............................101

Table 2-18. Correlation of PCs with behavioral measures.................................................................102
Table 2-19. Summary of PCA results........................................................................................................103

Table 3-1. Auditory Pupillometry Comparison 1 – Lexical Conditioning Experimental stimuli.................................................................128

Table 3-2. Auditory Pupillometry Comparison 2 – Structural Conditioning Experimental stimuli........................................................................129

Table 3-3. Non-variable governors used in Comparison 1– Lexical Conditioning.....132

Table 3-4. Negated variable governors used in Experiment 2 – Structural Conditioning..........................................................................................135

Table 3-5. Governors used in Filler items.........................................................................................137

Table 3-6. Splicing of experimental sentences – Auditory Pupillometry...............139

Table 3-7. Main model output – Lexical Conditioning.........................................................147

Table 3-8. Main model output – Structural Conditioning..............................................148

Table 3-9. Elicited Production – Lexical Conditioning Experimental stimuli........164

Table 3-10. Elicited Production – Structural Conditioning Experimental stimuli......165

Table 3-11. Cutting of experimental sentences – Elicited Production.........................166

Table 3-12. Elicited Production Model – Summary of Fixed Effects.........................170

Table 3-13. Participants’ self-reported characteristics.................................................174

Table 3-14. Self-reported proficiency (7-point Likert scale)...........................................174

Table 3-15. Diversity of language use across social contexts...............................175

Table 3-16. Main model output – Lexical Conditioning..............................................177

Table 3-17. Main model output – Structural Conditioning........................................178

Table 3-18. Elicited Production Model – Summary of Fixed Effects.........................183
LIST OF FIGURES

Figure 1-1. Subjunctive percentage by verb.................................................................34

Figure 1-2. Place of birth of Hispanics in New Mexico...............................................55

Figure 1-3. Bilingualism among Hispanics in New Mexico........................................62

Figure 2-1. Flowchart dimensions and their associated variables..............................76

Figure 2-2. Sample of question using percentage range slider..................................77

Figure 2-3. Sample of question reporting number of interlocutors..............................78

Figure 2-4. Sample of question reporting frequency of use per context.........................79

Figure 2-5. Sample of question reporting number of interlocutors per network group.................................................................81

Figure 2-6. Sample of question reporting bilingualism per network group..................82

Figure 2-7. Sample of question on attitude towards the heritage language..................83

Figure 2-8. Sample of question for identification rating...............................................84

Figure 2-9. Principal component projection simulation..............................................86

Figure 2-10. Contribution of variables to PCs – Heritage Language Background..........93

Figure 2-11. Contribution of variables to PCs – Contexts of Use.................................95

Figure 2-12. Contribution of variables to PCs – Productivity of Network.....................97

Figure 2-13. Contribution of variables to PCs – Bilingualism of Network....................98

Figure 2-14. Contribution of variables to PCs – Attitude, Motivation & Identity..........101

Figure 2-15. Correlation between PC1 – Contexts of Use and Spanish Picture Naming.........................................................................................................................108

Figure 2-16. Correlation between PC1 – Interactional Network and Picture Naming.........................................................................................................................110

Figure 2-17. Correlation between PC1 – Attitude, Motivation & Identity and Spanish Picture Naming.........................................................................................................................112

Figure 3-1. Stimulus distribution per list – Auditory Pupillometry.................................130
Figure 3-2. Person and Number distribution of Non-variable governors Comparison 1 - Lexical Conditioning

Figure 3-3. Person and Number distribution of subordinate verbs Comparison 1 - Lexical Conditioning

Figure 3-4. Person and Number distribution of negated governors Comparison 2 - Structural Conditioning

Figure 3-5. Person and Number distribution of subordinate verbs Comparison 2 - Structural Conditioning

Figure 3-6. Praat annotation scheme

Figure 3-7. Trial design – Auditory Pupillometry

Figure 3-8. Pupillary Response Smooths based on Fitted Values Lexical Conditioning – Non-variable Governors

Figure 3-9. Pupillary Response Smooths based on Fitted Values Structural Conditioning – Negated Governors

Figure 3-10. Pupillary Response Smooths based on Fitted Values Regularity – Non-variable Governors

Figure 3-11. Pupillary Response Smooths based on Fitted Values Regularity – Negated Governors

Figure 3-12. Difference in Pupillary Response based on Fitted Values Frequency – Non-variable Governors

Figure 3-13. Difference in Pupillary Response based on Fitted Values Frequency – Negated Governors

Figure 3-14. Difference in Pupillary Response based on Fitted Values Proficiency – Non-variable Governors

Figure 3-15. Difference in Pupillary Response based on Fitted Values Proficiency – Negated Governors

Figure 3-16. Difference in Pupillary Response based on Fitted Values Heritage Language Use – Non-variable Governors

Figure 3-17. Difference in Pupillary Response based on Fitted Values Heritage Language Use – Negated Governors
Figure 3-18. Difference in Pupillary Response based on Fitted Values Language Entropy – Non-variable Governors.................................................................159

Figure 3-19. Difference in Pupillary Response based on Fitted Values Language Entropy – Negated Governors.................................................................160

Figure 3-20. Difference in Pupillary Response based on Fitted Values Identification with the Heritage Language– Non-variable Governors..............................161

Figure 3-21. Difference in Pupillary Response based on Fitted Values Identification with the Heritage Language– Negated Governors..............................162

Figure 3-22. Stimulus distribution per list – Elicited Production............................166

Figure 3-23. Trial design – Elicited Production......................................................168

Figure 3-24. Difference in Subjunctive Production Accuracy by Condition............171

Figure 3-25. Difference in Subjunctive Production Accuracy by Condition, Regularity and Frequency.................................................................172

Figure 3-26. Pupillary Response Smooths based on Fitted Values Lexical Conditioning – Non-variable Governors.................................................................176

Figure 3-27. Pupillary Response Smooths based on Fitted Values Structural Conditioning – Negated Governors.................................................................177

Figure 3-28. Pupillary Response Smooths based on Fitted Values Regularity – Non-variable Governors.................................................................179

Figure 3-29. Pupillary Response Smooths based Fitted Values Regularity – Negated Governors.................................................................180

Figure 3-30. Difference in Pupillary Response based on Fitted Values Frequency – Non-variable Governors.................................................................181

Figure 3-31. Difference in Pupillary Response based on Fitted Values Frequency – Negated Governors.................................................................182

Figure 3-32. Difference in Subjunctive Production Accuracy by Condition............183
ACKNOWLEDGEMENTS

It is almost unbelievable that the time has come to write the acknowledgements section of this dissertation. Getting this far has required courage and determination on my part, but I could not have done it without the kindness and help that family, friends and colleagues have bestowed upon me throughout the years of this journey. First and foremost, I would like to thank my parents, Patricia and Edgardo, who have been my rocks over the past five years. They have always had kind—and tough—words of encouragement, especially on those days when I felt like giving up. I would most definitely not have reached my goal without your support, I owe and love you more than words can explain.

I must also thank all of my committee members: Dr. Paola E. Dussias, Dr. Matthew T. Carlson, Dr. Michael T. Putnam, Dr. Karen Miller and Dr. Damián Vergara Wilson, as well as my ex-advisor Dr. Cristóbal Lozano. Your role in shaping me as a scholar has been instrumental in the writing of this thesis. Over the years, you have seen me laugh and cry, develop into the researcher I am now, and stop doubting myself. Thank you to all of you for never stopping believing in me. In addition, it is of particular relevance to me to acknowledge the importance of other scholars’ work in my career. Dr. Dora LaCasse, Dr. Silvia Perez-Cortes and Dr. David Giancaspro, as you already know, your work has inspired me beyond belief. Your ideas and contributions have been a springboard for mine, you have all my gratitude for allowing me to build on your excellent work.

I would be remiss not to thank my fellow Graduate students, especially Dr. Michael A. Johns, without whose guidance, expertise and willingness to always help this dissertation would be much diminished. Mike, you have been my “crutch” over these
past five years and there is no way I can thank you enough for all the help you have given me. I would also like to thank Dr. Annie Beatty-Martínez, Dr. Isabel Deibel, Dr. Grant Berry, and Dr. Manuel Pulido Azpiroz for all their support. At different stages of my path, you have been wonderful friends and mentors in your own wonderful ways, for which I remain forever in your debt. Thank you also to Gabriela Puscama, your friendship is invaluable to me, as is your company in this scholarly journey of ours.

Finally, a big shout out to the state of New Mexico, the university of New Mexico and its scholars. None of this would have been possible if you had not opened the doors of your lab to me. New Mexico and its people have changed my life for the better, have given been a career, and love above all else. Along these lines, thank you to my fiancé Eduardo Orea for walking by my side in this difficult path, for putting up with the endless hours of writing and data analysis, for always cheering me up at the end of day.

My most sincere, heart-felt THANK YOU to everybody who has contributed to making this possible!

This research has also been supported by numerous grants and fellowships, both to me and others: The National Science Foundation Doctoral Dissertation Research Improvement Program (Grant No. BCS-1939903), and a National Science Foundation PIRE fellowship to Priscila López-Beltrán supported this research. I would also like to acknowledge National Science Foundation Grant Nos. 1019112 and 1019122 to Rena Torres Cacoullos and Catherine Travis, as well as Grant No. 1624966 to Rena Torres Cacoullos and Shana Poplack, which supported the creation of the New Mexico Spanish-English Bilingual corpus. Lastly, the research in this dissertation has been supported by
National Science Foundation Grants Nos. BCS-1535124, OISE-1545900, and the National Institutes of Health Grant Nos. HD082796 and HD071758 to Paola E. Dussias.
Para el Libro Gordo del Petete y La Gallega.

May you always watch over me.
Given that 59.25% of second language speakers in the U.S. are heritage speakers of Spanish (Nangano, 2015), it is not surprising that the linguistic knowledge of this population has been investigated and discussed extensively. Studies assessing the grammatical knowledge of speakers of Spanish as a heritage language, have primarily focused on the Spanish subjunctive mood. The subjunctive is a construction with a protracted acquisition pattern because it is largely infrequent in the input, its semantic connotations are abstract and its morphology is opaque, all of which complicates and delays its acquisition (Biber, Davies, Jones and Tracy-Ventura, 2006; Collentine, 2003). Given these characteristics, prescriptively mastering the Spanish subjunctive is considered a significant landmark of acquisition and an indicator of high proficiency.

The majority of studies examining heritage speakers’ knowledge of the subjunctive have concluded that their mastery of mood selection seems to be subject to phenomena such as incomplete acquisition, attrition or simplification, and that they present marked asymmetries in their interpretation and production of the indicative/subjunctive contrast (Montrul, 2007, 2009, et seq.; Silva-Corvalán 1994) Recently, however, an increasing number of studies have challenged these deficit-oriented perspectives, providing evidence that, while different from that of monolingual speakers, heritage speakers’ linguistic knowledge is by no means deficient (Bookhammer, 2013; Giancaspro, 2017; LaCasse, 2018; Perez-Cortes, 2016; Valdés-Kroff, Villegas & Dussias, under review). Given that there is a clear disconnect between previous research and recent findings, the present dissertation aims at bridging this gap by integrating corpus data into the design of psycholinguistic experiments that will examine heritage speakers’ and
Spanish-dominant bilinguals’ processing of the Spanish subjunctive during online comprehension and production. Critically, three methodological innovations are implemented. First, there is a dearth of studies conducted with long-standing bilingual communities; for this reason, the proposed research investigates knowledge of the Spanish subjunctive in heritage speakers from New Mexico, as well as a comparable sample of Spanish-dominant bilingual speakers from Mexico, to better understand how experience within a long-standing bilingual community modulates heritage speakers’ linguistic knowledge. Going beyond linguistic factors and considering the role of the community is important because linguistic variation is oftentimes modulated by social factors (Gumedstad, 2010, 2016, et seq; Labov, 1972). Second, most previous research has largely relied either on normative notions of mood derived from prescriptive grammars, anecdotal evidence or on small samples of naturalistic speech from first-generation immigrants to establish the factors constraining subjunctive selection. In the current experiment, all materials used were based on tokens extracted from a corpus of natural spoken language (Corpus Sociolingüístico de la Ciudad de México; Martín Butragueño & Lastra, 2011, 2012, 2015). When designing the materials, the main goal was to present participants with ecologically valid stimuli that mirrored natural utterances produced by speakers in their everyday lives; the original tokens were edited as little as possible in order to keep them true to real life usage while still having them conform to experimental rigorousness. Finally, previous research has focused mostly on production when it comes to morphosyntax (Polinsky, 2018) and has employed methodologies that are not entirely suitable for the study of heritage speakers. For instance, there is evidence that heritage speakers are usually at a disadvantage in written tasks (Montrul, Foote and Perpiñán, 2008). For this reason, this study uses, auditory pupillometry, a novel methodology within the field of heritage language research, to collect data on participants’ responses
to experimental linguistic manipulations of mood and the factors that constrain it. Given that this methodology is implicit and aural, it is more natural in relation to participants’ experience with the heritage language. In addition, a production experiment is also implemented to explore whether there are asymmetries between comprehension and production of mood selection in the current sample of participants (Perez, Cortes, 2016; Perez-Cortes, Putnam & Sanchez, 2019; Villegas, Demestre & Dussias, 2013) given that recent models of production connect production and comprehension (Chater, McCauley & Christiansen, 2016; MacDonald, 2013).

This dissertation also provides an evidence-based characterization of heritage speakers by using a sociolinguistic questionnaire in conjunction with Principal Component Analysis to examine which language-experience related factors best explain the variability inherent to this population. Given that the conditions under which the heritage language is acquired impact heritage grammar outcomes (Torres, 2019; see Aalberse & Muysken, 2013 for a review), taking into consideration not just proficiency, but also sociolinguistic factors will allow scholars to achieve a better understanding of heritage speakers’ knowledge of the Spanish subjunctive.

In all, by examining a novel community, using a novel methodology that incorporates corpus-based findings in the design of psycholinguistic experiments, and triangulating both comprehension and production experimental data, the present dissertation is poised to achieve a more holistic understanding of the nature of grammars of heritage speakers, and to contribute to a theory of language acquisition and processing in heritage language contexts that has greater explanatory adequacy.

**Approaches to the study of heritage grammars**
Even though heritage grammars have been extensively researched, consensus is still far from being reached on how to appropriately characterize them and their speakers. Understanding and explaining patterns of convergence and divergence between heritage and non-heritage grammars is undoubtedly one of the most challenging tasks faced by scholars within the field of heritage language research. In seeking to explain these patterns, a wide range of different theories have been proposed, which can be grouped into two broad, conceptual classes: (1) traditional approaches proposing that the underlying nature of heritage and non-heritage grammars is intrinsically different (Benmamoun, Montrul & Polinsky, 2010; Montrul, 2002, *et seq.*; Silva-Corvalán, 1994; Zapata, Sánchez, & Toribio 2005; *inter alia*), and (2) modern approaches, which rely on experience-based variables to explain differences and similarities between heritage and non-heritage grammars (Caroll, 2017; Pascual y Cabo, 2013; Pascual y Cabo & Rothman, 2012; Pascual y Cabo & Gómez Soler, 2015; Perez-Cortes et al., 2019; Pires & Rothman, 2009; Putnam & Sanchez, 2013; *inter alia*). While these approaches have all been extremely useful in understanding different aspects of heritage grammars, I argue that usage needs to take a central part in theories of language, as understanding the grammar of individuals as an outcome entails understanding the processes that shaped them (Putnam & Sanchez, 2013). Usage-based approaches, which have rarely been applied to the study of heritage grammars, place a strong emphasis on the effects of life-long experience with language use and, as such, they are an exceptionally well-suited tool to understand the process of heritage language acquisition (Bybee 2010, 2013; Goldberg, 2006; MacDonald, 2013).

In the following subsection, I will describe each of these approaches in detail, presenting their underlying rationales, as well as their relative strengths and weaknesses. It is important to note that the approaches discussed are neither mutually exclusive nor
inherently "correct" or "incorrect". While it is true that some of them have elicited more debate than others, achieving a unified theory of heritage language acquisition and processing requires a certain degree of common ground between all these approaches.

**Traditional approaches: incomplete acquisition and attrition**

The most prevalent traditional approaches are *incomplete acquisition* and *attrition*, which conceptualize heritage grammars as the product of divergent attainment. The incomplete acquisition approach is perhaps the most widely cited and, at the same time, the most controversial approach to the study of heritage grammars (Benmamoun, Montrul & Polinsky, 2010; Domínguez, Hicks & Slabakova, 2019; Montrul, 2002, et seq.; Polinsky, 2018; Silva-Corvalán, 1994; Zapata et al., 2005). Under this perspective, divergence between heritage and non-heritage grammars implies differences in underlying knowledge that are reflected in phenomena such as grammatical reduction, simplification and reanalysis (for reviews see Benmamoun, Montrul & Polinsky, 2013; Polinsky, 2018; Polinsky & Scontras, 2019). It has been argued that because heritage speakers experience a reduction in the quantity and quality of input in their L1 (minority language) at a very early age they may acquire mental representations of grammatical structures that are fundamentally different from those of non-heritage speakers. For instance, Domínguez et al., (2019) explained that “incomplete acquisition can arise when input conditions change during the course of acquisition (and, crucially, before a mature and stable grammar has fully developed) and that change impacts upon the overall pattern of acquisition and resulting grammar” (p. 242). Thus, incomplete acquisition is the result of arrested development influenced by a dominance shift toward the L2 (majority language).
As an alternative, the attrition approach argues that heritage grammars have undergone a process of attrition (Köpke 2007; Schmid & Köpke, 2013). The extent to which L1 skills are subject to attrition is directly correlated to the age of onset of bilingualism (Flores, 2010; Montrul, 2008, 2011, 2016; Pallier, 2007). Studies on attrition have shown that the earlier heritage speakers are introduced to the L2, the more different their resulting grammars are from those of monolingual speakers (Merino, 1983; Montrul, 2016; Polinsky, 2008, 2011, 2016, 2018; Silva-Corvalán, 2003, 2013). However, it is, in fact, not clear whether incomplete acquisition can be differentiated from attrition, since there is no way to, *a posteriori*, prove whether a particular construction was never acquired (incomplete) or acquired and then eroded (attrited) (Putnam & Sanchez, 2013).

While the work by Montrul and Polinsky greatly advanced the field of heritage language research in the early 2000s, divergent attainment approaches have been heavily criticized, especially by sociolinguists. In the case of the incomplete acquisition approach, the main argument against it is that the term “incomplete” suggests that the process that monolinguals follow is a “normal” trajectory, culminating in a fully and successfully developed grammar; by comparison, heritage speakers are supposed to follow a deviant course leading to an unsuccessful outcome (Bayram, Kupisch, Pascual y Cabo & Rothman, 2019; Felser, 2020; Flores & Rinke 2019; Kupisch & Rothman, 2018; Otheguy, 2013; Pascual y Cabo & Rothman, 2012; Putnam & Sánchez, 2013; Putnam, & Kupisch, 2018; Putnam, Carlson & Reitter, 2018; *inter alia*). Thus, not only does this entail that the process of acquiring a heritage language is inherently deficient, but this concept is grounded on the notion that there is an idealized grammatical system homogenously reached by all monolingual speakers that represents a *complete* linguistic target to be acquired.
Characterizing a linguistic system as complete entails two major dilemmas that have not been addressed so far: (1) how to define and (2) how to quantify completeness. Regarding definition, it seems that complete equates a prescriptive standard language spoken by native speaker population that shares an identical mental grammar. However, there is ample evidence that there is variation in how native speakers represent their L1 (see work by Dąbrowska, 2008A, 2008B, 2012; Dąbrowska & Street, 2006; Street & Dąbrowska, 2010; also see Sorace & Keller, 2005 for a review). Concerning quantification, it has been argued that heritage grammars are primarily the result of input reduction, which implies that there must be a quantifiable amount of input necessary to achieve a complete grammatical system. The incomplete acquisition approach fails to provide information on how to measure this amount of input, as well as to define what amount would be crucial. In addition to this, detractors of a view of heritage grammars as the product of attrition have argued that this approach has traditionally posited a view of language that is too linear and under which linguistic knowledge can only be preserved or lost (but see de Leeuw, Opitz & Lubinski, 2013; Schmid, Köpke & de Bot, 2013; Seton & Schmid, 2016). However, there is abundant evidence that linguistic knowledge changes dynamically throughout our lives (Birdsong, 2018; Chang, 2012; Dussias & Sagarra, 2007; Seton & Schmid, 2016), that even receptive bilinguals can possess detailed knowledge of the morphosyntax of a grammar (Sherkina-Lieber 2011, 2015), and that element of grammar that seem to have been lost can be revitalized (Dahl et al., 2010; Pierce et al., 2014).
Current approaches: input quality and lexicalist/activation

Given the evidence that changes in the quantity of input do not preclude language loss (Flores, 2015; Hudson-Kam & Newport, 2005; Putnam & Sanchez, 2013), recent approaches have sought to go beyond the impact of fluctuations in input quantity to account for divergence between heritage and non-heritage grammars. Importantly, these approaches have considered more experience-based factors, taking into consideration the properties of the input heritage speakers are exposed to, as well their usage of the heritage language.

Under the input quality approach, divergence can be explained by examining how the language knowledge of heritage speakers is shaped by differences in the input they receive from their parents (Dominguez, 2009; Flores, 2015; Montrul & Sánchez Walker, 2013; Pascual y Cabo, 2013; Pascual y Cabo & Rothman, 2012; Pascual y Cabo & Gómez Soler, 2015; Pires & Rothman, 2009; Prada Perez & Pascual y Cabo, 2011; Rothman, 2009; Salazar, 2007; Silva-Corvalán, 1994). In other words, if some of the properties of the input on which heritage grammars are built are already different from—or more variable than—established prescriptive norms, then, it cannot be expected that heritage speakers will effortlessly process or produce features that conform to norms they have not been exposed to. A good example of this case can be found in a study by Pires and Rothman (2009) who examined heritage speakers’ knowledge of inflected infinitives in Brazilian Portuguese (BP). Inflected infinitives are a formal feature of BP which is learned almost exclusively in educational settings. Because the heritage speakers in their sample had had no formal exposure to BP, they neither produced nor understood inflected infinitives in comparison to educated BP speakers, who learned and mastered inflected infinitives in school. Thus, the authors explain that their sample of heritage speakers did not use this construction not because they lacked knowledge, but because they did not have sufficient
exposure to grammars that instantiated the higher-register properties of inflected infinitives robustly. In other words, speakers cannot be expected to produce constructions they do not have experience with (see also Dominguez, 2009; Rothman & Treffers-Daller, 2014).

However, it is uncommon for non-heritage grammars to exhibit properties that are entirely absent in their heritage counterparts. More often, what is observed is that first-generation immigrants (i.e., heritage speakers’ parents) begin to exhibit emerging variability with certain properties of the heritage language, and heritage speakers either replicate or amplify and extend variable patterns present in their input (Flores, 2015; Montrul & Sánchez Walker, 2013; Prada Perez & Pascual y Cabo, 2011; Silva-Corvalán, 1994). For instance, Pascual y Cabo (2013) investigated heritage speakers’ knowledge of the reverse psychological predicate gustar (‘to like’); because gustar is non-agentive, it does not appear in passive constructions, in comparison to other psychological predicates such as, for example, molestar (‘to annoy). The results of an acceptability judgment task suggested that heritage speakers showed a tendency to accept gustar in passive constructions, a trend which is argued to have originated in the emerging instability of the predicate gustar in the grammatical systems of first-generation immigrants. While first-generation immigrants in the study did not show a strong tendency to accept gustar in passive constructions, Pascual y Cabo posited that even minor variability in the first-generation immigrants’ use of gustar could have been greatly amplified in heritage grammars (see also Bookhamer, 2013; Otheguy, 2013).

Nevertheless, this approach alone cannot account for all patterns of divergence. Heritage grammars sometimes show divergence even in constructions that are invariant and frequent in the input. For instance, heritage speakers sometimes produce and accept gender-mismatched Spanish noun phrases (Cuza & Pérez-Tattam, 2016; Montrul, 2016;
Montrul, Foote & Perpiñán, 2008; Montrul & Potowski, 2007). The activation approach (Putnam & Sanchez, 2013) proposes that age of acquisition, input quantity and input quality alone cannot be the only deterministic factors explaining the differences observed between heritage non-heritage grammars. From this perspective, divergence stems from differences in the strength of activation of the heritage language; in other words, the frequency with which heritage speakers use the target language for production and comprehension modulates its activation and, therefore, the ease with which they access it. This approach is relevant and innovative because it brings to attention how usage shapes heritage grammars. Under this account, variability stems from speakers’ relative use of the heritage language, which makes it easier or more difficult to access and produce certain forms, thus giving rise to differences not only with respect to non-heritage grammars, but also between heritage grammars. Of particular interest to activation approach is the fact that production/comprehension asymmetries are oftentimes observed in heritage populations (Mikulski, 2006; Mikulski & Elola, 2013; Montrul, 2009, 2011; Putnam & Sanchez, 2013; Rothman, Pascual y Cabo & Lingwall, 2012; Sherkina-Lieber, Pérez-Leroux & Johns, 2011). For instance, Perez-Cortes (2016) examined heritage speakers’ knowledge of the Spanish subjunctive and showed that, except for those speakers whose proficiency level was very advanced, all participants were more accurate in tasks that required the interpretation of subjunctive, than those that required production. The author argued that heritage speakers do have complete knowledge of the subjunctive, but that their access to their L1 knowledge for production is especially impacted.
Applying usage-based approaches to the study of heritage grammars

While usage-based approaches to language have a great potential to advance our understanding of heritage grammars, they have rarely been used within the field of heritage language studies (to my knowledge, only Akkus, 2019; Karayayla, 2020; LaCasse, 2018).

Within linguistics, the question of how experience with language influences the cognitive representations of speakers’ grammars first became relevant in the 1980s (Du Bois, 1985; Givón, 1979; Hopper & Thompson, 1984). So called usage-based approaches to language propose that, rather than being a pre-requisite for language use, language knowledge emerges from usage. Language structure is shaped by the every-day interaction of language use and domain-general cognitive mechanisms. These mechanisms modulate the linguistic choices that speakers make during production which, over time, generate distributional patterns in the language input that comprehenders are exposed to. In turn, these distributional patterns become the probabilistic constraints that guide comprehension (Bybee 2008, 2010, 2013; Goldberg 2006; MacDonald, 2013).

Within these approaches, the main unit of representation is the construction; these are form-meaning pairings that exist at multiple levels of linguistic abstraction (Bybee, 2006; Ellis, 2008, et seq.; Goldberg, 2006) and are learned throughout individuals’ lives as they identify the distributional regularities in their input. In addition, usage-based approaches emphasize that linguistic knowledge is also shaped by social factors because it is “grounded in and emergent from language use in concrete social activity for specific purposes that are tied to specific communities of practice” (Hall, Cheng & Carlson, 2006: 235). Thus, constructions are social configurations, shaped by the socially-situated contexts within which they are acquired.
While incompleteness and attrition do acknowledge the importance of experience, they give it a very limited role. On the contrary, usage-based approaches to language offer a nuanced account of the relationship between experience and acquisition, detailing how experience dynamically shapes usage and processing. This view shifts the focus away from whether there is sufficient experience for a given structure to be acquired—or to prevent its attrition—asking instead what structure we can expect to emerge given different amounts and kinds of experience, and how changes in the frequency of specific structures can lead to a gradual evolution of grammatical patterns over time. Thus, it offers a principled way of understanding the origins of what might, on the surface, appear to be incomplete acquisition or attrition (Putnam & Sanchez 2013; Rothman 2009; Flores 2015; Yang 2016).

Recapitulation
Heritage grammars have been studied from multiple approaches, all of which make a robust and valid contribution to the field of heritage language studies. However, despite the fact that heritage grammars have been extensively researched consensus is still far from being reached, approaches that take into account the crucial role experience and usage-based factors have the potential to provide a more unified and accurate account of language acquisition and processing in heritage contexts. In the following section of this chapter, I provide an overview of important findings within the field of heritage language research.
The linguistic knowledge of heritage speakers

Throughout the years, a vast amount of evidence has been accumulated showing that heritage speakers diverge from other speaker groups in systematic ways. Studies within the field of heritage language research have concentrated on elucidating which domains of heritage languages are resilient (i.e., more native-like), and which are vulnerable (i.e., prone to attrition or incomplete acquisition). Two overarching trends are reported in the literature: (1) heritage speakers’ phonological systems are more native-like than their knowledge of morphosyntax (for reviews see Montrul, 2015; Polinsky, 2018; Polinsky & Scontras, 2019); and (2) evidence suggests that heritage speakers are better at oral tasks as well as in tasks that are implicit and require less metalinguistic knowledge (Alarcón, 2011; Montrul, Foote & Perpiñán, 2008; Montrul, 2010). Despite these trends in the extant literature, there is also evidence to the contrary, showing that heritage speakers can converge with monolingual speakers at all linguistic levels. In the following subsections, I provide an overview of the most relevant findings within the field of heritage language research, reviewing and discussing studies that compare heritage speakers both with monolinguals and L2 learners. Importantly, given that the literature covering the linguistic knowledge of heritage speakers is vast, I concentrate on illustrating crucial points using only a select number of structures; however, additional citations referring to similar results with other structures are also provided.

Findings on phonology

While heritage speakers are quite close to monolingual speakers of the heritage language both in production and perception, they still exhibit certain differences in their knowledge of the phonetic and phonological systems of the heritage language (Au,
An across-the-board finding is that heritage speakers have a “global accent”, that is, to monolinguals their speech sounds foreign both in the heritage and non-heritage languages. In a seminal study, Yeni-Komshian, Flege and Liu (2000) recorded heritage speakers (L1 Korean–L2 English) as they repeated sentences in Korean (all of them grammatical). A control group of monolingual Korean speakers listened to the recordings and were asked to evaluate whether the recorded participants had a foreign accent. Results showed that monolinguals rated heritage speakers as having a distinctive non-native accent when speaking Korean (for similar results see: Bae, 2015; Godson, 2003, 2004). Similar results have been found for heritage speakers’ dominant language (i.e., their L2); a study by Kupisch et al., (2014a) examined L1 French–L2 German and L1 Italian–L2 German heritage speakers’ production. A group of monolingual speakers of German was asked to evaluate the accent of heritage speakers as native or foreign in recordings of naturally produced speech. It was found that, although monolinguals identified heritage speakers as native speakers of their dominant language (i.e., German), they systematically noted that they sounded “foreign” (for similar results see Flege 2007; Flege et al. 1999).

The perception of heritage speakers’ accent as foreign can be traced, at least in part, to the fact that their languages influence each other in subtle ways given that the interaction between any two languages is bidirectional (Polinsky, 2018). For example, in Spanish the lateral approximant /l/ is described as “clear” while in American English /l/ is characterized as “dark”. While the distribution of Spanish /l/ is quite consistent across contexts, “its American English counterpart, however, is subject to an allophonic
velarization rule: [l] (dark /l/) occurs in the syllable rhyme (e.g., meal [mil], candle [kændl]), while [l] occurs in the syllable onset (e.g., lease [lis], fly [flaɪ]) (Polinsky, 2018: 139). L1 Spanish–L2 English heritage speakers acquire allophonic variation of laterals in English and Spanish relatively early, such that their velarization patterns are target-like in both languages. However, studies by Barlow, Branson and Nip (2013) and Barlow (2014) show that there is a cross-linguistic influence from Spanish in their production of English lateral approximants, such that regardless of the context [l] appears in, it is pronounced as if it were clear (for similar results see Jannedy & Weirich, 2014). Barlow and colleagues hypothesized that heritage speakers may thus ignore differences where their two languages are relatively similar, which inevitably leads to a certain foreign accent.

While heritage speakers’ production differs from that of monolingual speakers in nuanced ways, heritage speakers seem to be quite close to monolinguals in perception. In a seminal study, Tees and Werker (1984) studied whether being exposed to a second language in the early years of acquisition was correlated with a facilitating effect in the domain of perception. One of the populations they studied were English-speaking adults who had been exposed to Hindi as infants. Results showed that even at low proficiency levels, these heritage speakers maintained the ability to discriminate between phonemic contrasts to the same extent as a group of native Hindi controls, suggesting that early exposure to Hindi phonemic contrasts influenced their discriminatory ability later in life (Polinsky, 2018: 153).

Along these lines, it is a common finding that heritage speakers seem to have an advantage over L2 learners within phonology and phonetics (Godson, 2004; Gor, 2014; Lukyanchenko and Gor, 2011; Oh et al., 2010; Simpson & Cooke, 2005). For instance, Au
et al., (2002) and Knightly et al., (2003) investigated the pronunciation of Spanish heritage speakers compared to L2 learners. They conducted a series of phonetic analyses in order to assess participants’ accents, examining voice onset time, degree of lenition and phoneme production. Results revealed that heritage speakers’ accent was more native-like than that of L2 learners (see also Au et al., 2008). Along the same lines, Chang, Yao, Haynes and Rhodes (2011) examined the production of back vowels, aspirated plosives, and retroflex fricatives in Mandarin in heritage speakers, L2 learners and native speakers. Once again, they found that heritage speakers outperformed learners, being able to perceive and produce sounds more accurately. Similarly, Saadah (2011) examined vowel production in heritage speakers of Palestinian Arabic and L2 learners. In line with previous studies, the results suggested that heritage speakers’ pronunciation was more native-like than that of L2 learners (for similar results with other phenomena see Kerswill, 2002; Nützel & Salmons, 2011; Larsson, Tingsell & Andréasson 2015; Lein, Kupisch & van de Weijer, 2016; Otheguy, Zentella & Livert 2007).

Thus, it seems that access to two distinct phonologies makes heritage speakers better at phonological categorization and perception of contrasts, suggesting an advantage to being an early bilingual. However, age alone cannot be the most deterministic factor explaining the differences observed (Pascual y Cabo & Rothman, 2012). Alternatively, the results of the studies discussed so far, and the so-called advantage of heritage speakers over L2 learners, can also be explained by considering each population’s distinct language acquisition experience. First, while learners acquire their L2 in a classroom setting, engaging in mostly written interactions with their instructors, many heritage speakers acquire their L1 in a naturalistic setting, engaging in conversations with their caregivers and members of their community. Thus, any observed advantages on phonology are a product of their highly oral linguistic experience.
It must be taken into consideration, however, that there are heritage speakers who also have instructional experience with the heritage language, especially those living in places where the heritage language is prevalent, as in, for instance, New Mexico.

In all, the evidence presented so far suggests that at the phonological and phonetic levels, heritage speakers can indeed present native-like qualities. However, it is in perception rather than production that they are more like monolingual speakers of the heritage language. In addition, the studies reviewed in this subsection make it clear that it is difficult to merely categorize heritage speakers’ linguistic knowledge as native-like or not; rather, one must consider a myriad of aspects such as the skill being tested, the linguistic phenomenon, the language dyad, proficiency, and age of acquisition, among others, before being able to draw a conclusion. This is robust proof that a “one-fits-all” explanation of heritage-speakers’ abilities holds little validity.

**Findings on morphosyntax**

In contrast with the studies reviewed so far, research has shown that heritage speakers’ production and interpretation of morphological and syntactic structures shows more, and stronger, divergence from monolingual speakers than their phonological systems (for a comprehensive review see Polinsky, 2018, chapter 5). Studies suggest that the *irregularity* of morphosyntactic paradigms determines, to a large extent, how these are acquired by heritage speakers: overall, elements that are more irregular and less perceptually salient are more difficult to acquire. For instance, Korean is famous for its register or speech-level contrasts which are instantiated via different endings in finite verbs for each register. There are six distinct registers determined by the speaker’s relationship to the addressee: deferential, plain, polite, blunt, familiar, and intimate.
A study by Choi (2003) showed that heritage speakers tended to either eliminate or transform irregular stem alternations across registers. Their repair strategy relied on reanalysis of the form available in the intimate register, which is the one heritage speakers are most familiar with, as they receive most of their Korean input from their families. Thus, it is to be expected that heritage speakers will retain more strongly those morphosyntactic structures and elements which they have learned in those contexts where the heritage language is majorly used.

Another aspect that modulates heritage speakers’ acquisition of morphosyntactic phenomena is salience. Given the importance of perceptual salience to heritage comprehension and production, it is not surprising that heritage speakers do not do well with “silent” elements. Morphosyntactic structures such as relativization strategies, form-mappings and the interpretation of null pronouns are some silent aspects of heritage syntax that seem to be particularly prone to non-convergence. For instance, Polinsky (2011) investigated the comprehension of relative clauses in Russian heritage speakers and Russian monolinguals using an auditory picture-matching task. Crucially, she manipulated not just the type of relative clause (object vs. subject), but also the word order of these clauses (subject-before verb vs. verb-before subject). Results showed that while monolinguals performed at ceiling in all structures, heritage speakers did well with subject relatives but performed at chance on object relatives with both word orders. Polinsky explained that “heritage speakers showed a divergent grammar of relativization [because] they associated a gap in the relative clause with the missing subject of that clause and did not seem to entertain the possibility that this gap could also correspond to another clause constituent” (2011: 35), (see also: Sánchez-Walker, 2012; for production see Albirini & Benmamoun, 2014).
Similar results are found in studies examining the interpretation of null pronouns (Laleko & Polinsky, 2017; Polinsky, 2018; Polinsky & Scontras, 2019; see also Ivanova-Sullivan, 2014; Montrul, 2004; Pérez-Leroux & Glass, 1999; Rothman, 2009; Serratrice, Sorace & Paoli, 2004; Sorace & Filiaci, 2006; Tsimpli, 2007; \textit{inter alia}; see Keating, VanPatten & Jegerski 2011 and Polinsky, 2018, for in depth reviews of this topic). It has been argued that because heritage speakers exhibit a preference for one-to-one mapping between form and function, this entails that they have difficulty in producing and interpreting linguistic segments that contain null elements such as anaphoric traces (for similar convergence results with other constructions see Albirini, Benmamoun & Saadah, 2011; Cuza & Frank, 2011; Hoot, 2017; Johannessen & Laake, 2015; Mai, 2015; Scontras, Degen & Goodman, 2017; Westergaard, 2009).

Another common divergence pattern is that there is a general preference in heritage grammars for one-to-one mappings between form and meaning (Laleko & Polinsky, 2017; Polinsky, 2018), a trend that is especially salient in languages with case marking. In many of these languages, the dative can serve multiple functions, resulting in the mapping of a single case into multiple syntactic positions (for a review see Irizarri van Suchtelen, 2014). Studies have shown that this multiple mapping is simplified in heritage grammars, typically by restricting the dative form to the role of the indirect-object recipient, which is its canonical role (Montrul, Bhatt & Girju, 2015; Polinsky 2006, 2018).

In addition to the evidence presented in support of divergence so far, studies comparing heritage speakers’ and L2 learners’ knowledge of morphosyntax have not yielded clear-cut results on the so-called \textit{advantage} of heritage speakers over their L2 learner peers. In fact, most studies have found that the performance of heritage speakers and L2 learners is largely comparable in this domain (Håkansson, 1995; Mikhaylova,
2012; Montrul, Bhatt & Girju, 2015; Montrul & Perpiñán, 2011; Putnam & Salmons, 2015).
For instance, a study of verb movement in Spanish conducted by Bruhn de Garavito (2002), and another on case markers and relative clauses in Korean by O’Grady, Lee, and Choo (2001) revealed no differences between heritage speakers and L2 learners, in other words, they diverged from the monolingual speaker baseline in the same way, such that their interpretation of object relative clauses was more inaccurate, and they oversimplified case markings. Within derivational morphology, a significant number of studies has focused on investigating gender and case systems to test whether they are simplified in heritage grammars. In a seminal study, Polinsky (2008) examined heritage speakers’, L2 learners’ and monolingual speakers’ knowledge of gender and case in Russian. Results suggested that both heritage speakers and L2 learners had very simplified case declensional systems which, in turn, affected their knowledge of gender assignment, especially in the case of low frequency nouns. Likewise, Montrul and Bowles (2009) investigated the knowledge of the accusative and dative marking preposition *a* in Spanish—also called Differential Object Marking (DOM; e.g., *María vio a Juan*, ‘Mary saw John’; *María dio un beso a Juan*, ‘Mary gave John a kiss’) heritage speakers and L2 learners at different proficiency levels, as well as native speakers. Their results indicated that heritage speakers did not significantly differ from L2 learners. In addition, even at advanced levels of proficiency, heritage speakers’ use of DOM was probabilistic and erratic, and their knowledge unstable even in the most frequent cases, such that, for example, ungrammatical constructions sometimes had the same acceptance rate as grammatical ones (for similar results see Montrul, 2002; Montrul & Sánchez-Walker, 2013; Montrul, Bhatt, & Girju, 2015; Zapata, Sánchez & Toribio, 2005). It has been argued that the dative marking preposition *a* in Spanish “is one of the best-documented victims of imperceptibility in heritage speech” (Polinsky, 2018: 166) because the lack of perceptual
salience on this marker a makes it especially vulnerable for heritage speakers and L2 learners alike.

The literature discussed so far illustrates instances of divergence between heritage speakers and other speaker groups. However, there is also a growing body of literature showing that convergence at the morphosyntactic level is indeed a possible outcome. For example, Prada Pérez (2009) examined heritage and non-heritage speakers of Spanish and Catalan, and found no difference in the rate of overt subjects between speaker-groups. Likewise, Keating et al., (2011, 2016) investigated heritage speakers of Spanish processing null versus overt pronouns during online comprehension, in comparison to a monolingual baseline. They examined heritage speakers’ knowledge of impersonal null pronouns, as well as their knowledge of the contrast between referential null and overt pronouns and concluded that the performance of both groups was not significantly different (for similar results see Dubinina & Polinsky, 2013; Laleko & Polinsky, 2017).

As explained before, some studies have reported that heritage speakers have difficulties processing anaphoric dependencies because connecting linguistic material at a distance in the heritage language (i.e., non-dominant language) taxes their working memory (for reviews see: Montrul, 2016; Polinsky, 2006, 2016, 2018). However, there is also evidence that while anaphora resolution may indeed be difficult for heritage speakers, they do have knowledge of the lexical properties of anaphoric pronouns. For instance, Korean has a local anaphor, caki-casin, and a long-distance anaphor, caki, (Cole, Hermon & Huang, 2001). While here are a number of constraints that govern the use of these pronouns, only the constraint that the antecedent of caki must be a third person animate is categorical (Choi & Kim, 2007; Kang, 1998). It has been hypothesized that heritage speakers would have more difficulty with the constraints of caki that are variable, rather than with the one that is categorical. Nevertheless, comprehension studies of
English-dominant heritage speakers of Korean have shown that these speakers have native-like knowledge of both the variable and categorical lexical constraints of *caki* and *caki-casin* (Kim, Montrul & Yoon, 2010, 2015). Thus, the difference between monolingual and heritage speakers in the interpretation of anaphoric dependencies is quantitative rather than qualitative: both speaker groups possess robust knowledge of this syntactic phenomenon, but heritage speakers have difficulty recovering antecedents at a distance (for similar convergence results with other constructions see Guasti, 2017; Hulk & Müller, 2000; Kim & Goodall, 2016; Putnam & Salmons, 2013).

In all, this subsection sums up results that are like those of the previous one: while, at the morphosyntactic level, heritage speakers tend to show systematic divergence from monolingual speakers, they can also exhibit native-like knowledge in some constructions. Thus, once again, a clear-cut asseveration that heritage speakers’ linguistic knowledge is fundamentally non-native like is incorrect.

**Task effects**

Ample evidence has been provided that heritage speakers are overall better at online oral tasks as well as in tasks that are implicit and require less metalinguistic knowledge. For instance, Montrul, Foote and Perpiñán (2008) studied knowledge of grammatical gender in Spanish in heritage speakers and L2 learners at different proficiency levels, as well as monolingual speakers. Results showed that, although neither group had native-like mastery of gender in Spanish, heritage speakers’ answers were more accurate in oral tasks (i.e., picture description) than in written tasks (i.e., gender completion task and written gender recognition task; see Alarcón, 2011 for similar results). Likewise, Montrul (2010) examined the knowledge of various forms of object expression in Spanish (i.e., clitics, clitic left dislocations, and differential object marking) in heritage speakers, L2
learners and monolingual speakers as well. Once again, it was reported that neither bilingual group had native-like knowledge of object expression and that both were affected similarly by transfer from English. However, heritage speakers performed better than L2 learners in oral tasks, producing overall more clitic pronouns. In addition, it was found that although heritage speakers’ performance was not as good as that of L2 learners in written tasks, they showed higher acceptance rates of sentences with object topicalizations, which are more common in spoken than written language (see also Méndez, Rothman & Slabakova, 2015).

Concerning explicitness, Bowles (2011) administered a battery of tests to heritage speakers, L2 learners and monolingual speakers to measure implicit and explicit knowledge of several aspects of Spanish morphosyntax (i.e., ser/estar “be,” gender, a personal, preterit/imperfect, subjunctive, adjective placement, conditionals, and subject-verb agreement). Results indicated that heritage speakers performed better in tasks such as interpretation tasks, oral narratives, and timed grammaticality judgments—which reflect implicit knowledge—than in morphology recognition tasks or untimed grammaticality judgments—which tap into more explicit knowledge. A potential problem with Bowles’ study is that it confounded explicitness and modality. For this reason, Montrul, Davidson, De la Fuente and Foote (2014) conducted a similar study in which they implemented three spoken word recognition experiments that varied on the degree of explicitness of the task: a gender monitoring task, a grammaticality judgment task, and a repetition task. Their aim was to test how the implicitness or explicitness of these tasks interacted with heritage speakers’, L2 learners’ and monolinguals’ knowledge of grammatical gender in Spanish. According to their results, all groups showed sensitivity to gender violations in Spanish, but heritage speakers outperformed learners (i.e., were more native-like) in the implicit of tasks (i.e., the repetition task).
These findings might be explained by considering that the acquisitional differences between heritage speakers and other speaker groups are compounded by the fact that many heritage speakers have low literacy skills in the heritage language throughout their lives, receiving mostly auditory input and using this language in the oral modality for the most part. This entails that, oftentimes, they do not have exposure to the written form of the heritage language and do not develop their metalinguistic skills to the same extent as L2 learners do. It is not surprising, then, that they have less difficulties with tasks that are oral and/or implicit. As mentioned before, however, it must be noted that heritage speakers’ experience with the heritage language in instructional settings is very varied and many heritage speakers receive formal education in the heritage language both during their teenage years and during adulthood.

Recapitulation
The studies reviewed throughout this section have highlighted the fact that convergence between heritage speakers and other speaker groups is possible. While a growing body of experimental work has focused on documenting deficits in the heritage grammar and exploring the differences between heritage and baseline versions of the same language, it is now clear that divergence is not an inevitable outcome. Divergence and convergence are complex phenomena, guided by a myriad of linguistic and experiential factors. These phenomena stem not only from differences in age of acquisition, quality/quantity of input or activation patterns. Individuals’ socially-situated experience with language shapes not just their linguistic knowledge, but the way they process language both for comprehension and production. Importantly, rather than an indication that they are better
or more native-like than any other bilingual group, differences are just natural reflections of distinct experiences with language acquisition and use.

In the following section, I provide a detailed characterization of the construction under study in this dissertation, the Spanish subjunctive. Here, I thoroughly review both the traditional (i.e., semantic and syntactic) and variationist accounts of mood selection in Spanish. Importantly, this next section establishes a parallelism with the current section as it highlights how it is by empirically examining the usage of the subjunctive that the variables that condition mood selection in Spanish have been uncovered and purely prescriptivist accounts have been debunked.

The subjunctive mood

While Spanish can be considered to have three moods: indicative, subjunctive and imperative (Seco, 1973), it is the subjunctive that has overwhelmingly elicited interest among scholars. The subjunctive is a grammatical mood which consists of a series of morphological inflections used to mark modality (Bosque, 2012). Most of these inflections are formed by a shift in the thematic vowel—in bold here: indicative 3rd sg. = quiere vs. subjunctive 3rd sg. = quiera, ‘to want.’ However, the subjunctive can also be marked by both a shift in thematic vowel and a change in the verbal root: indicative 3rd sg. = traer vs. subjunctive 3rd sg. = traiga, ‘to bring.’ Regarding its syntactic distribution, while the subjunctive can appear in main clauses following certain adverbials (e.g., ojalá ‘hopefully,’ quién ‘if only,’ quizás, acaso, tal vez, ‘maybe’ (see Butt & Benjamin, 2000), it occurs primarily in subordinate clauses (i.e., relative, complement, adjectival, and adverbial clauses, see Dozier & Iguina, 2003).

The characterization of the Spanish subjunctive—much like in other Romance languages—has been the subject of ample debate within the linguistic community.
Traditionally, prescriptive studies on the nature of mood in Spanish have proposed either semantic or syntactic explanations of subjunctive selection (Ahearn & Leonetti, 2004; Gómez-Veiga, García-Madruga & Moreno-Ríos, 2010; Heras Sedano, 2006; Palmer, 2001; Quer, 1997, 2009; Villalta, 2008; inter alia), and a few more recent theories have tried to provide models that integrate both semantic and syntactic notions (Bergen, 1978; Bosque, 2012; Fábegras, 2014; Farkas, 1992; Foster, 1973; Kempchinsky, 1995, 2009; Quer, 2009). Nevertheless, not only is there still no consensus on a theoretical characterization of the subjunctive, but these normative definitions can readily be contested with examples from corpora of natural language. Modern variationist studies based on corpora of natural spoken language have provided evidence indicating that the Spanish subjunctive is highly lexically routinized; in other words, subjunctive selection is primarily conditioned by lexical factors, with structural (i.e., syntactic) and semantic elements playing only a minor role (Torres Cacoullos, LaCasse, Johns & De La Rosa Yacomelo, 2017; Poplack, Torres Cacoullos, Dion, Berlinck, Digesto, LaCasse & Steuck, 2018).

Throughout this chapter, I will first consider traditional research on the origins and nature of the Spanish subjunctive, as a brief overview of the extant literature on the origins of mood in Spanish will be useful in understanding the foundations of the assumptions and biases underlying the traditional characterizations of the subjunctive that are currently prevalent. After this, I will discuss in detail how modern variationist studies have empirically challenged traditional characterizations of mood selection. Finally, I will conclude with a subsection devoted to an in-depth review of the research on heritage speakers’ knowledge of the Spanish subjunctive within the context of the US.
Origins of the subjunctive

The Spanish subjunctive has its origins in a common Romance ancestor: the Latin subjunctive, which is said to have developed as a merge of the subjunctive and optative moods of Proto-Indo-European. Some scholars have proposed that the meanings of the subjunctive were potential and volitive, while the optative conveyed fictive and wishing meanings (Bennett, 1910). Thus, present-day subjunctive is oftentimes defined as the mood of potential, tentative, hypothetical, ideal, or even unreal action (irrealis) in contrast to the indicative, which is the mood of actuality and factuality (realis; see Wheelock, 2000). However, alternatives to this semantic evolution have also been proposed: Bybee, Perkins and Pagliuca (1994) argued that while the Latin subjunctive was initially used in main clauses in which it had independent meaning, its use eventually extended to complement clauses in which its meaning harmonized with that of the main clause; this modal redundancy caused the semantic properties of the subjunctive to be gradually bleached, until it ceased to provide additional meaning to the sentence, which enabled it to continue to expand to clauses that were not modally harmonic. Through this grammaticalization process, the subjunctive eventually became reanalyzed as a mere concomitant of subordination, becoming a syntactic reflex rather than providing semantic contribution (see also Magni, 2010).

These plausible but contradictory explanations highlight the fact that, to this day, there is disagreement among scholars about the purported origins and meanings of the subjunctive. For instance, Murphy (2008) writes that “after centuries of scholarship […] opposing viewpoints are the norm rather than the exception with regard to the subjunctive mood in Romance languages” (pp. 5-7).
Traditional characterizations of the subjunctive

Traditionally, the nature of the Spanish subjunctive has been described from a semantic and syntactic perspective (for an in-depth review see Murphy, 2008). Semantically, the indicative/subjunctive distinction has most commonly been explained using the semantic contrast between realis/irrealis (Bolinger, 1974; Bosque & Demonte, 1999; Escarpenter, 1974; Farkas, 1992; Portner & Rubinstein, 2012; Quer, 2009; Terrell & Hooper, 1974; Whitley, 2002; inter alia). The concept of realis (1.1. a) refers to factual situations that are occurring at the moment of the utterance, whereas irrealis (1.1. b) denotes situations from the realm of hypotheses and alternative worlds (Haverkate, 2002; Palmer, 2001):

(1.1)

a. **Realis**: La doctora cree que estás\textsubscript{IND} enfermo
   ‘The doctor believes you are\textsubscript{IND} ill.’

b. **Irrealis**: La doctora quiere que vayas\textsubscript{SUBJ} a casa.
   ‘The doctor wants you to go\textsubscript{SUBJ} home.’

Because of its connection with the realis/irrealis distinction, mood selection has also been associated to the concepts of contextual commitment, veridicality, and presupposition. Under this approach, the choice between indicative and subjunctive is a reflection of the speaker’s commitment towards the truth-value of the embedded proposition. For instance, in (1.1. b) the subjunctive is used because the possibility exists that the patient will not go home no matter what the doctor says, which makes the presupposition non-assertive or irrealis. On the contrary, (1.1. a) introduces first-hand knowledge information about the truth-value of the statement, making it assertive or realis.
As an alternative to semantic accounts, there is a rich body of literature which supports that the nature of the subjunctive can be better explained by integrating both syntactic and semantic notions (Bergen, 1978; Bosque, 2012; Fábegas, 2014; Farkas, 1992; Foster, 1973; Kempchinsky, 1995, 2009). The most widespread account is that of Quer (2009), who proposed a distinction between intensional and polarity subjunctive. Intensional subjunctive is selected by a lexical element (e.g., verb, preposition, or conjunction) and requires a specific type of mood, forbidding any possible alternations. For instance, in example 1.b above, a sentence in the indicative such as La doctora quiere que *vas.IND a casa, would be considered ungrammatical, as the matrix verb requires the subjunctive. Polarity subjunctive appears in predicates where the indicative is also possible when licensed by a structural element in the sentence (e.g., negation, interrogation) or by a specific semantic need (e.g., presupposition). In example 1.a, the sentence La doctora cree que estás.IND enfermo, the indicative is used to assert the existence of a non-presupposed sickness, because she is certain the patient is ill.

This dichotomy between uses of the subjunctive that are governed by syntactic vs. semantic constraints is mirrored in the traditional division between obligatory vs. optional contexts of use, which is predominantly found in the literature examining heritage speakers’ knowledge of the subjunctive (Gutierrez, 1994; Lozano, 1974; Lynch, 1999; Marroco, 1974; Martínez Mira, 2009; Montrul, 2007, 2009, et seq.; Ocampo, 1990; Silva-Corvalán, 1991, 1994, et seq.; van Osch & Sleeman, 2018; inter alia). Under this division, contexts that are considered obligatory select the subjunctive categorically and are usually conditioned by a lexical element, whereas within optional contexts, the subjunctive and indicative can alternate, constrained by a myriad of semantic and structural factors.
The sheer number of accounts proposed to characterize mood selection creates an intrinsic complication for scholars because when studies use different categorizations in an idiosyncratic manner, comparison across experiments and the extrapolation of their results becomes challenging. To illustrate this, Table 1.1 below shows the variability in the definition of what constitutes obligatory contexts in three studies, Viner (2016), Silva-Corvalán (1994) and Montrul (2009). The study by Viner is the most comprehensive, examining a total of 8 contexts, followed by Silva-Corvalán who examined 6 contexts; while there is some overlap between the first two studies such that three of the obligatory contexts investigated are the same (i.e., purpose, causative and hypothetical clauses), a comparison with Montrul’s work cannot even be established given the underspecification of the contexts under study.

Table 1.1 Variability in definition of obligatory contexts of use across studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Contexts</th>
</tr>
</thead>
</table>
| **Viner (2016)** | • Discourse marker (*o sea, ‘I mean’).  
• Purpose/Contingency adverbial clause (*para que canten, ‘so they sing’).  
• Causative clause: *hace que canten / he makes them sing’).  
• Indirect command (*dijo que cantaran / he told them to sing’).  
• Temporal adverbial clause with futurity (*cuando canten, ‘when they sing’).  
• Imperative (*canten, ‘sing!’).  
• Protasis in hypothetical conditional clause (*si cantaran, ‘if they sang’).  
• Hypothetical *como si clause (*como si cantaran, ‘as if they sang’). |
| **Silva-Corvalán (1994)** | • Volitional and causative matrices  
• Matrices with expressions of impossibility.  
• Hypothetical manner clauses.  
• Temporal clauses introduced by a conjunction (*antes, ‘before’).  
• Purpose clauses (*para que, ‘so that’)  
• Comment matrices (*es mejor que, ‘it is better if’) |
| **Montrul (2009)** | • Contexts where the subjunctive is subcategorized by the verb or subordinating expression. |
This case in point exemplifies a common problem among studies researching the subjunctive, underscoring the need for a more unified account of mood selection. In the following subsection, I discuss a recent variationist study that used corpora of natural spoken language to empirically uncover the constraints conditioning mood selection in Spanish.

**Variationist characterizations of the subjunctive**

While normative characterizations of mood are largely prevalent in the literature, there is in fact scant empirical evidence to support the extent to which the Spanish subjunctive is governed by the myriad of semantic, structural, and pragmatic functions that have been attributed to it by both grammarians and linguists (LaCasse, 2018: 13). Formalist notions of mood such as “polarity vs. intensional subjunctive” and “obligatory vs. optional contexts of use” have a number of shortcomings. First, normative characterizations based on the purported semantic or pragmatic functions of the subjunctive can readily be contested with examples from corpora of natural language. To illustrate, it is prescriptively assumed that negated evaluative predicates (e.g., *no creer*, ‘to believe-NEG’) denote disbelief on the part of the speaker when used with the subjunctive (*Yo no creo que sea-SUBJ inocente*), while they express the speaker’s firm belief with the indicative (*Yo no creo que es-IND inocente*). On these grounds, it has been argued that the subjunctive is the only grammatical option to convey non-presupposition with first- and second-person singular matrix subjects (Quer, 1999; Kempchinsky, 2009, 2016). However, a conversational corpus of Spanish (CORLEC, Marcos Marín, 1992) and historical Spanish texts (e.g., *La Celestina*)

---

1 Translation: ‘I do not believe s/he is subj/ind innocent.’
both show variation between indicative and subjunctive in these contexts, as can be seen in the following example from the CORLEC (Torres Cacoullos et al., 2017: 2-3):

(1.2) Pero... no creas que allí se comportan,\textsubscript{IND} igual. ‘Don’t think\textsubscript{2nd sg.} they behave\textsubscript{IND} the same way there.’

Second, because prescriptive accounts are not based on the observation of natural language use, they fail to explain variation patterns that break away from their prescriptive norms and do not consider the constrains that condition mood selection in real life language usage. For instance, they cannot account for the fact that there is variation in subjunctive selection among monolingual varieties of Spanish (Gallego & Alonso-Marks, 2014; Kowal, 2007; Lastra & Burtragueño, 2012; Serrano, 1992; Pousada, Poplack & Sankoff, 1982). In addition, a diachronic, Pan-Romance corpus study of mood selection by Poplack et al., (2018), which investigated both early and modern texts, has shown that the Romance subjunctive has undergone a process of desemanticization (Haspelmath, 1998; Heine & Reh, 1984) and conventionalization (DeMulder & Lamiroy, 2012; Lehmann, 2002); this entails that the association between the linguistic context and mood selection has been increasingly strengthened over the years, leading to an almost total loss of purely semantically or syntactically motivated choice between indicative and subjunctive.

Drawing from this study, Torres Cacoullos et al., (2017) employed a quantitative variationist approach to empirically investigate the factors that condition usage of the Spanish subjunctive across time. The authors analyzed the subjunctive in complement clauses in three major works within Spanish literature, ranging from the XIII to the XVI centuries, comparing them to modern Spanish in the Corpus Sociolingüístico de la Ciudad de México (C SCM; Martín Butragueño & Lastra, 2015). Torres Cacoullos and colleagues
examined five factors conditioning mood selection: (1) type frequency, (2) proportion of the data accounted for by the subjunctive, (3) the subjunctive rate of each matrix verb, (4) polarity and (5) sentence type. Results showed that while the subjunctive was more productive in early Spanish, it is now highly lexically conditioned (i.e., routinized) such that, in modern Spanish, a high number of infrequent verbs are the ones that favor subjunctive selection the most in both periods. Importantly, these verbs present a bimodal distribution (Figure 1.1), that is, most verbs are Non-variable (71%), selecting either the subjunctive or the indicative almost categorically, while only a few verbs show variability; in fact, it was reported that within the 23 most frequent verbs, only five were robustly variable (between 22% and 78%). The fact that most verbs are Non-variable stresses the fact that subjunctive selection in Spanish is primarily conditioned by the identity of the verb, as only a minority of verbs allowed for influence of semantic or structural factors. Among these subjunctive governing verbs (governors henceforth) volitional, influence and causative verbs are identified as the “bulwark” of the Spanish subjunctive (Torres Cacoullos et al., 2017: 17); governors such as querer, ‘to want’, desear, ‘desire’ or rogar, ‘beg’ select the subjunctive categorically; in comparison, evaluative and/or epistemic verbs such as puede ser, ‘it might be that’, pensar, ‘to think’ or dudar, ‘to doubt’ present more variable subjunctive selection rates.
In addition to the evidence presented so far, LaCasse and Johns (2018) investigated the prescriptive premise that the semantic class of the governor is a primary motivator for subjunctive selection. Their results suggested that the idiosyncratic rates of subjunctive selection are not driven by the semantic class of any given governor, but rather that the most frequent individual governors within each semantic class are behind the apparent semantic effect (p.56; for similar results see Poplack, Lealess & Dion, 2013).

Regarding the non-lexical factors that influence mood selection, Torres Cacoullos et al. examined polarity and sentence type. Results indicated that negative, interrogative, and conditional clauses favored subjunctive selection; for instance, see the effect of polarity in Table 1.2 below, where it is shown that the subjunctive selection rate was higher with negated governors both in early (71%) and modern Spanish (80%). In addition, the effect of negation was stronger for variable governors in both periods as well (57% and 69% respectively).
Table 1-2. Effect of non-lexical factors on mood selection

<table>
<thead>
<tr>
<th>Polarity</th>
<th>Early Spanish</th>
<th>Modern Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All governors</td>
<td>Variable governors</td>
</tr>
<tr>
<td>negative</td>
<td>71% (117/166)</td>
<td>57% (64/113)</td>
</tr>
<tr>
<td>affirmative</td>
<td>57% (489/888)</td>
<td>31% (170/547)</td>
</tr>
</tbody>
</table>

Torres Cacoullos et al. concluded that the Spanish subjunctive is highly lexically routinized, as mood selection is first and foremost conditioned by a series of matrix verbs, with some minor effect of local structural elements, and almost no effect of semantic import. These results, based on natural spoken data, debunk previous prescriptivist theories which explain subjunctive selection based solely on semantic and syntactic factors.

Acquisition of the subjunctive

Studies on the acquisition of the Spanish subjunctive in monolingual populations have shown that this construction has a protracted developmental schedule, such that although contexts in which the subjunctive is selected almost categorically are acquired around 3 years of age, children do not have full, adult-like mastery of all the nuances and uses of the subjunctive until after the age of 8 years (Blake, 1983; Pérez-Leroux, 1998; Sánchez-Naranjo & Pérez-Leroux, 2006). One of the reasons behind its protracted acquisition is the fact that the subjunctive is largely infrequent in the input. Biber, Davies, Jones and Tracy-Ventura (2006) examined the frequency of over two-million verbs in the Corpus del Español (Davies, 2000) and found that the subjunctive only accounted for 7.2% of all verbs forms. In addition, because the distribution of the subjunctive is largely restricted to more formal, written registers, its overall frequency in speakers’ input is
further decreased (Collentine, 2003). Another two factors that make acquiring the subjunctive challenging are that its semantic connotations are abstract, and its morphology is opaque whenever the verb is regular (e.g., regular: *quiero* - 1st sg. Ind. vs. *quiera* - 1st sg. Subj., ‘to want’; irregular: *voy* - 1st sg. Ind. vs. *vaya* - 1st sg. Subj., ‘to go’); because of this, the subjunctive is oftentimes not salient to speakers (Collentine, 2003).

Studies with monolingual and bilingual\(^2\) populations have shown that mood selection is first learned in “obligatory” contexts of use, (i.e., intensional subjunctive, deontic predicates, Non-variable governors) and then in “optional” contexts of use (i.e polarity subjunctive, epistemic/epistemological predicates, variable governors; Blake, 1983; Merino, 1983; Lozano, 1995; Silva-Corvalán, 2014). However, the acquisition of mood differs for monolinguals and bilinguals. In monolingual populations, the pioneering work by Ana Teresa Perez-Leroux laid the foundations of mood acquisition theory. In her 1998 seminal study, she proposed that maturational constraints mediate the acquisition of the subjunctive in different contexts, arguing that children need to be cognitively ready to handle situations regarding speaker’s presuppositions and beliefs in order to acquire the full spectrum of mood selection nuances (also see de Villiers & de Villiers, 2000). For instance, mood selection is first acquired with deontic predicates because these only necessitate comprehending meanings such as obligation and permission; by comparison, the implicatures of epistemic/epistemological predicates are more difficult to conceptualize because they require the consideration of the speaker’s point of view and their comments of the validity of the statement. Thus, the understanding of false beliefs (irrealis) is a cognitive prerequisite for the acquisition of the evaluative values of the subjunctive. This hypothesis has been supported by numerous

\(^2\) By bilingual populations, I refer to those relevant to the present dissertation, namely, heritage speakers of Spanish, or L1 English-L2 Spanish learners.
studies showing that children are more accurate in their production interpretation and use of mood in obligatory contexts of use than in optional contexts that require the computation of complex semantic import (Blake, 1983; Echeverría, 1978; Gallo-Valdivieso, 1994; Padilla, 1990; *inter alia*).

In comparison to the L1 literature, there are only a few studies that examine the acquisition of mood selection bilingual children (Anderson, 1999; 2001; Cuevas de Jesús, 2011; Merino, 1983; Perez-Cortes, 2016; Silva-Corvalán, 2003, 2014). Overall, while studies report the same developmental trends in bilingual and monolingual children, they also emphasize that there are variables at play in bilingual environments that modulate the acquisition of mood. Frequency of use of the minority language (i.e., Spanish) at home, and the variety of people and contexts in which children use the language are the two factors which exert the most influence in mood acquisition. For instance, Silva-Corvalán (2014), who analyzed the bilingual development of her two grandsons, documented that the youngest received less exposure to Spanish and, therefore, showed a weaker command of mood, tending substitute the subjunctive with infinitives, imperatives, and present indicative forms. While studies of mood selection in bilingual children are scarce, the subjunctive has been thoroughly studied in adult bilingual populations. In all, most of the extant literature has shown that both heritage and L2 speakers of Spanish retain the use of subjunctive morphology mostly in obligatory contexts, while they tend to be more variable in optional contexts. In the following section, I provide a summary of studies dedicated to the acquisition and processing of the Spanish subjunctive by adult heritage speakers. In addition, I also discuss the literature comparing knowledge of mood in heritage speakers and L2 learners of Spanish; while L2 learners are not the topic of the present dissertation, comparisons between them and heritage speakers have often been
made and, as such, are relevant to achieve a better understanding of the acquisition of mood in bilingual contexts.

**Bilingual grammars and the Spanish subjunctive**

*Heritage speakers*

Heritage speakers’ knowledge of the Spanish subjunctive has been studied from multiple theoretical frameworks. Among these, sociolinguistic/variationist (Anderson, 2001; Bookhamer, 2013; Guitart, 1982; Lantolf, 1978; Lozano, 1974; Lynch, 1999; Viner, 2017; Merino, 1983; Ocampo, 1990; Pousada & Poplack, 1979; Puente-Schubeck, 1991; Silva Corvalán, 1994, 2001, *et seq.*; Waltermire, 2014; Zentella, 1997; *inter alia*) and linguistic studies are, by far, the most prevalent (Giancaspro, 2017, 2019; Iverson, Kempchisky & Rothman, 2008; Montrul, 2004, 2007, *et seq.*; LaCasse, 2018; Montrul & Perpiñán, 2011; Pascual y Cabo, Lingwall, & Rothman, 2012; Perez Cortes, 2016; Requena, Dracos & Miller, 2017; van Osch & Sleeman, 2018; *inter alia*). In general, studies dedicated to the analysis of mood selection in adult heritage speakers document significant differences between heritage and non-heritage speakers’ (i.e., monolingual speakers) knowledge of the subjunctive. A recurrent pattern that has been reported across studies is that subjunctive selection in “optional” contexts, where mood alternation is—allegedly—semantically relevant, is the most problematic for heritage speakers, while knowledge of “obligatory” contexts, where the subjunctive is lexically-selected, is more native-like and less prone to simplification. Based on these differences, the mood systems of heritage speakers have oftentimes been described as being *reduced, simplified, attrited* or *incomplete*

---

3 Also see pedagogical (Mikulski, 2006) and cognitive research (Potowski, Jegerski & Morgan Short, 2009).
Within the variationist field of research, the work by Silva-Corvalán (1994, 2001, 2010) is the most widely cited. Her seminal 1994 study focused on the gradual, inter-generational loss of the indicative/subjunctive alternation in the Spanish grammar of bilinguals in Los Angeles, California; her goal was to understand how contact with English throughout time might have affected heritage speakers’ use of the subjunctive in obligatory and optional contexts. In this study, Silva-Corvalán interviewed three groups of speakers within this community: the first group included first-generation, Mexico-born immigrants living in the US, the second group included second-generation, US-born heritage speakers whose parents had been born in Mexico, and, finally, the third group included third-generation, US-born heritage speakers who had at least one US-born parent. Results showed a progressive decrease of subjunctive use in all contexts across generations, such that third-generation speakers’ rates of subjunctive use were the lowest; the reduction was weaker in volitional, causative, purpose, and concessive clauses, all of which were considered obligatory contexts of use, while optional contexts were more prone to being simplified and replaced by the indicative. In addition, it was found that reduction in subjunctive use was correlated with the fact that proficiency in Spanish also decreased in each generation. Silva-Corvalán concluded that the community under study in her work was undergoing and inter-generational process of simplification of mood. Given that subjunctive loss was already noticeable in the first generation, she concluded that this process was accelerated by contact with English (for similar results see Anderson, 2001; Guitart, 1982; Lantolf, 1978; Lozano, 1974; Lynch, 1999, 2000; Martínez Mira, 2006; Merino, 1983; Ocampo, 1990; Silva Corvalán, 1994, et seq.; Zentella, 1997).
Among the linguistic studies, Silvina Montrul, whose work is couched within the Generative framework (Chomsky, 1986), was a pioneer in the experimental study of heritage speakers of Spanish knowledge of mood. (Montrul, 2004, 2007, 2009, 2010, 2011, 2013, 2016, 2018; Montrul & Perpiñán, 2011). Montrul (2007) examined how heritage speakers at different proficiency levels interpreted the subjunctive in comparison to non-heritage speakers; her sample consisted of 20 heritage speakers from different backgrounds (mostly from Mexico, Perú, Colombia and Puerto Rico), and 15 monolingually-raised native speakers from different Spanish-speaking countries (p. 29). Participants completed a morphology recognition task designed to test their knowledge of mood in obligatory contexts, and a sentence conjunction judgement task to test their knowledge of mood in optional contexts. The results of the morphology recognition task indicated that while monolinguals performed at ceiling, heritage speakers were overall less accurate, such that only those at advanced proficiency levels displayed moderately robust knowledge of subjunctive in obligatory contexts. According to the results of the sentence conjunction judgement task, only monolinguals showed clearly differentiated interpretations of subjunctive vs. indicative use in optional contexts; although advanced heritage speakers were more native-like, even at intermediate levels of proficiency, participants failed to show robust knowledge of mood distinction. While there was a clear effect of proficiency, Montrul concluded that heritage speakers seemed to be unable to discriminate semantically between subjunctive and indicative in optional contexts, where there is a subtle meaning difference (2007: 37). From her theoretical perspective, this is caused by the loss of a functional category that carries the morpho-phonology and semantic features necessary for mood distinction.

In a follow-up study, Montrul (2009) recruited a bigger and more homogeneous sample (23 native speakers and 65 Spanish heritage speakers that were born in the US to
Mexican parents) and employed more varied tasks (oral elicitation task, morphology recognition task and sentence conjunction task) to re-examine heritage speakers’ knowledge of mood in obligatory and optional contexts. Once again, heritage speakers differed significantly from native speakers, especially in the tasks testing optional contexts, but there was a clear developmental trend such that the higher the proficiency level, the more native-like heritage speakers’ knowledge of mood was. For instance, while heritage speakers at advanced proficiency levels were the only ones to discriminate between indicative and subjunctive in the sentence conjunction task, they did so to a much lesser extent than the non-heritage baseline; similarly, all heritage speaker groups tended to over-produce indicative forms in the elicited production task, and only heritage speakers produced mood errors in this task, even at advanced levels of proficiency. Montrul argues that because the subjunctive is a complex structure that is acquired relatively late, this makes it particularly vulnerable in heritage speakers whose input in the L1 is drastically reduced early in life. Thus, according to her, the results observed suggest that heritage speakers’ knowledge of mood in Spanish is affected by incomplete acquisition during childhood (for similar results to Montrul, 2009, see van Osch & Sleeman, 2018).

In conclusion, the studies on the acquisition of Spanish mood discussed so far revealed proficiency to be a determining factor in the accurate use of subjunctive and indicative forms. In addition to this variable, heritage speakers’ knowledge also seems to be modulated by the obligatory/optional nature of the mood selection context under investigation, being more native-like in constructions where the subjunctive is lexically selected (i.e., obligatory contexts, intensional subjunctive) than in those where mood selection is determined by semantic/pragmatic factors (i.e., optional contexts, polarity subjunctive).
While the majority of the literature within the field of heritage language research has reported findings similar to the ones above, increasingly studies report results that challenge these deficit-oriented and reductionist portrayals of heritage speakers’ knowledge of mood. Within the variationist paradigm, Bookhamer (2013) used sociolinguistic interviews to examine the use of mood in first (i.e., immigrants) and second-generation (i.e., US born) heritage speakers of Spanish representing the six main Spanish-speaking groups in New York City: Puerto Rican, Dominican, Mexican, Ecuadorian, Colombian and Cuban. His goal was to investigate whether second-generation heritage speakers’ grammars had changed under contact with English. In order to do so, he did not rely on a prescriptive obligatory vs. optional contexts division; rather, he used the production of first-generation speakers to construct a baseline of contexts which selected the subjunctive categorically or variably; he then compared second-generation speakers’ production to this baseline. Bookhamer found that second-generation heritage speakers produced subjunctive forms in all the same contexts as first-generation speakers, albeit at a lower frequency and with more within-group variability (see Viner, 2016, for more detailed results). Thus, the author concluded that rather than referring to the grammar of second-generation heritage speakers as having undergone a process of simplification, loss or incomplete acquisition, it might be more accurate to define it as being variable; under this definition, a group of speakers’ usage of mood can be uniform in some contexts, and more variable in others, but still be overall highly systematic and consistent. Importantly, he emphasized that despite differences in the use of mood by both groups of speakers, all participants achieved successful, nuanced communication; in fact, he notes that if both groups are analyzed independently, they display “unique, yet equally complete and systematic, mood grammars” (p. 97).
Linguistic studies have also provided evidence against the simplification of the mood system in heritage grammars. For instance, Pascual y Cabo, Rothman and Lingwall (2012) examined the performance of 47 heritage speakers with different levels of proficiency in obligatory (i.e., desideratives) vs. optional (i.e., negated epistemics) mood selection contexts using a morphological multiple-choice task and a felicitousness judgment task. Results showed that accuracy rates on subjunctive selection in obligatory contexts was very high and appeared to be correlated with heritage speakers’ proficiency. While heritage speakers’ scores were lower in the judgment task, when their performance was compared to that of the non-heritage controls results indicated target-like acquisition of subjunctive even in optional contexts. Similarly, Perez-Cortes (2016) tested 137 heritage speakers on their knowledge of the intensional subjunctive with querer que (‘want that’) and the polarity subjunctive with decir que (‘say that’) using four experimental tasks: a truth-value judgment, two production tasks (written and oral), and an acceptability judgment task; she found that heritage speakers’ rate of subjunctive production was nearly identical with polarity and intensional subjunctive, which she argues is due to the fact that she controlled for propositional modality in her experiment (i.e., all stimuli were deontic predicates). In addition, there were proficiency effects in her data, such that except for those speakers whose proficiency level was very advanced, all participants were more accurate in those tasks that required the interpretation of subjunctive, than those that required production. Along these lines, Giancaspro (2017) investigated heritage speakers’ knowledge of intensional subjunctive mood with para que (‘so that’) and polarity subjunctive mood in adjectival relative clauses. He tested 42 heritage speakers at different proficiency levels and with different ages of acquisition, comparing them to 20 Spanish-dominant bilinguals who had acquired English early or late in life. Results showed that heritage speakers produced the subjunctive across the same contexts.
as their baseline and that their use of this construction was constrained by the same linguistic factors. The groups only differed in the frequency with which the subjunctive was used such that heritage speakers produced fewer subjunctive forms. In addition, this study sought to better understand which factors modulate heritage speakers’ knowledge of mood selection. Importantly, Giancaspro found that the earlier heritage speakers had acquired English, the more variable their knowledge of the subjunctive was. Moreover, knowledge of subjunctive selection with infrequent verbs also tended to be more variable regardless of age of acquisition (for similar results see: Giancaspro, 2019, and Iverson, et al., 2008). Both Perez-Cortes and Giancaspro used traditional off-line methods, however, there exists the possibility that using online, psycholinguistic methods may provide further insights into the richness of heritage speakers’ linguistic knowledge. Valdés Kroff, Villegas and Dussias (under review) explore this hypothesis using eye-tracking to examine knowledge of the subjunctive in heritage speakers of Spanish, in comparison to a Spanish-dominant group of L1 Spanish-L2 Catalan bilinguals. The authors investigated processing of the mood selection in ambiguous sentences where the governor (i.e., main verb) subcategorized for the subjunctive, thus acting as the disambiguating element (e.g., *El general les aconsejó a los soldados que trajeran_\text{SUBJ} \sim \text{trand}_\text{IND} la comida de casa [...]*). Results indicated that although heritage speakers had overall slower reading times than the Spanish-dominant participants, both groups showed an effect of mood such that verbs that appeared in the indicative morphology were read more slowly than those in subjunctive, as predicted if participants were able to access the verb subcategorization information encoded in governor. Thus, this study provides further evidence that

---

4 Translation: ‘The general advised the soldiers to bring the food from home.’
heritage speakers can have knowledge of mood similar to that of non-heritage speakers and can process the subjunctive in native-like ways even during online comprehension.

The literature discussed throughout this section has underscored the fact that (1) using a wide array of methodological approaches (2) leaving behind prescriptive norms to investigate what linguistic factors actually modulate mood selection in natural language use is crucial, not just to provide an explanatorily adequate characterization of subjunctive use, but also to achieve a better understanding of heritage speakers knowledge of mood, moving away from reductionist, deficit-oriented approaches. Given the population of interest in this dissertation, it is critical to review the work by LaCasse (2018), which has provided robust evidence against the inevitability of contact-induced simplification in the mood systems of heritage communities. This variationist study examined use of subjunctive in complement clauses using the New Mexico Spanish-English Bilingual Corpus (NMSEB; Torres Cacoullos & Travis, 2018). Most sociolinguistic studies investigating heritage speakers’ knowledge of mood selection have based their claims of simplification on the observation that heritage speakers produce overall lower rates of subjunctive in comparison to a set baseline. However, overall rates can be misleading because “statistically different rates need not translate into grammatically relevant differences” (Torres Cacoullos & Travis, 2015: 82; see also: Gallego & Alonso-Marks, 2014; Pousada & Poplack, 1979). Moreover, overall rates are often established as benchmarks based on prescriptive notions derived from reference grammars, which entails that they are not truly informative about the actual linguistic conditioning of variants (LaCasse, 2018: 61; see also Blake, 1983 for further empirical evidence that prescriptive benchmarks do not reflect real language use).

LaCasse abstained from using overall rates, employing instead the variationist method to explore the differences in the conditioning of the subjunctive and the indicative
in the NMSEB, using the *Corpus Sociolingüístico de la Ciudad de México* (CSCM; Martín Butragueño, 2011, 2012, 2015) as a usage-based benchmark rooted in natural production, and which therefore incorporates the inherent variation of language (p. 63). Her results showed no evidence of simplification in the grammars of New Mexican heritage speakers: speakers in both corpora used the subjunctive with comparable frequency and under the same linguistic constraints. LaCasse also reports that 78% of governors in the NMSEB were variable (vs. 39% in the CSCM), which indicates that there is no loss of variable contexts of subjunctive use in this contact variety of Spanish. In addition, more detailed measures of lexical routinization (i.e., proportion of subjunctive by governor frequency–frequent vs. non-frequent–and proportion of subjunctive by governor variability–variable vs. Non-variable) further reinforce the parallel patterns of variation in the NMSEB and CSCM, providing no evidence for accelerated change (p. 82). These results raise the possibility that what other studies have called *simplification* solely based on overall rates of subjunctive use, might not be so if the subjunctive is actually used across the *same contexts* by all populations, albeit to *different extents*.

Another critical finding by LaCasse, was that English did not affect subjunctive selection in her sample. Her analyses suggest that heritage speakers were extending the subjunctive to new contexts, by using Spanish subordinate clauses under English governors; seven English lexical governors occurred at least once with the subjunctive in Spanish, as in (1.2) below (2018: 114):

(1.2) *I wished que alguien lo pudiera.SUBJ agarrar.*
*‘I wished that somebody would grab him.’*

The fact that these results differ so greatly from previous literature highlights not just possible methodological issues across studies, but also the fact that the concept of
community has been largely disregarded, especially in linguistic studies. A heritage speaker’s community is crucial because the influence of the social environment on L1 acquisition is established from a very early age (Nardy, 2008). Lacasse’s work underscores the fact that not all bilingual communities are the same; New Mexico, for one, is the longest established bilingual community in the US, so is not surprising that the minority language has remained vital. On the contrary, studies like that of Bookhamer (2013) examine much more heterogeneous and “young” bilingual communities, such as that of New York City, and find much more variable speakers. Likewise, many other studies do not even concentrate on particular communities but recruit heritage speakers from varied backgrounds at academic institutions (e.g., Montrul, 2007, 2009, et seq.).

Going beyond linguistic factors and considering the role of the community is also important because linguistic variation is oftentimes modulated by social factors (Labov, 1972, 1984). In the case of the Spanish subjunctive in particular, social conditioning has rarely been reported; as Gudmestad points out, “existing research does not provide conclusive evidence on the role of social variables” in mood selection (2010: 41). To my knowledge, there are only two studies report such effects and both happen to examine the Spanish of New Mexico: Puente-Schubeck (1991) studied loss of subjunctive in New Mexican Spanish; she found that variables such as growing up in a Spanish-only household, self-reported comfort speaking Spanish, and membership in the two oldest generations favored subjunctive selection. Similarly, Waltermire (2014) examined social conditioning of the subjunctive in the speech of Spanish-English bilinguals from the Barelas neighborhood of Albuquerque. While he investigated age, gender, and language preference in his analysis, only language preference was selected as significant in his statistical analyses.
**L2 learners**

Knowledge of mood selection has elicited much interest among L2 scholars as well; while L2 learners are not the topic of the present dissertation, comparisons between them and heritage speakers have often been made and, as such, are relevant to achieve a better understanding of the acquisition of mood in bilingual contexts.

L2 Spanish learners’ use of mood morphology has been examined in a wide range of contexts (adverbial clauses: Ahern, Amenos-Pons & Guijarro Fuentes, 2014; Gudmestad, 2012, 2013, 2014; Kanwit & Geeslin, 2014; adjectival constructions: Borgonovo, Bruhn de Garavito & Prévost, 2008; Restorik Elordi, 2012; and nominal clauses: Iverson, Kempchinsky & Rothman, 2008; Lubbers-Quessada, 1998; Mikulski, 2006; Mikulski & Elola, 2013; Massery & Fuentes, 2014; *inter alia*). While most research has focused on learners’ production of indicative and subjunctive in obligatory and variable contexts (Collentine, 1995), a few studies have also examined comprehension (Cameron, 2011).

Before delving into this literature, it is important to first consider differences between mood in Spanish and English to better understand the learning task that L2 Spanish speakers undertake. Briefly, the main differences between the Spanish and English subjunctive are found in the distribution and morphology of mood in each language\(^5\). First, the subjunctive is much less frequent and productive in English and is only restricted to formal, written registers (Kleiser, 2008; Kovács, 2009; Ojea, 2005; Palmer, 2001); this entails that learners’ exposure to this construction is low in their L1, which can make them less sensitive to mood cues in the L2 as well. Second, Spanish and English grammaticalize mood differently (López Sancio, 2014); morphologically speaking, use of

\(^5\) For an in-depth study comparing the subjunctive in Spanish and English, the reader is referred to López Sancio (2014).
the subjunctive mood is only detectable in English in third person singular verb forms, as it prescriptively involves the dropping of the final –s, as in *he eats*–3rd sg. IND vs. *he eat*–3rd sg. SUBJ. While all other forms of the subjunctive are identical to the indicative in English, an exception to this rule is the verb *to be*, which in the subjunctive is simply expressed as *be* in all forms as in *I require that you be*–2nd sg. SUBJ *here now* (Kleiser, 2008; López Sancio, 2014).

The most productive use of the English subjunctive used to be in mandative constructions (Hundt, 1998; Crawford, 2009; Ojea, 2005; Övergaard, 1995; Serpollet, 2001), as in (1.3) below:

(1.3) *I request that he finish*–3rd sg. SUBJ *the job properly.*

However, by means of analogical extension, these subjunctive forms have been increasingly displaced by non-finite constructions (1.4 a), prepositional complementizers heading over by non-finite constructions (1.4 b), or modal verbs conveying necessity or obligation (1.4 c), (Ojea, 2005: Palmer, 2001):

(1.4)

a. *I request him to finish*–Infinitive *the job properly.*

b. *I request for him to finish*–Infinitive *the job properly.*

c. *I request that he should*–Modal *finish the job properly.*

Thus, the L1 English–L2 Spanish learner is faced with the task of detecting sites of morphological and structural conflict between both languages, and internalizing the constraints mood selection in the L2; for instance, L2 speakers must learn that:
- The subjunctive in Spanish requires the complementizer *que* (‘that’).
- Subjunctive verbs had distinct morphology, which might be regular or irregular.
- It is the governor that primarily conditions subjunctive selection, rather than a preposition.

Finally, a certain degree of crosslinguistic transfer might be expected in mandative constructions given that these are most structurally similar in Spanish and English (Kempchinsky, 2009; Ojea, 2005; Quer, 2009; Radford, 2007; Roussou, 2009) by virtue of including the complementizer and a verb with distinctive morphology (I request that he finish - 3rd sg. SUBJ the job properly → *Yo solicito que haga* - 3rd sg. SUBJ *el trabajo correctamente*). Thus, it might be possible to hypothesize that volitional *that*-subjunctive clauses (which fall into the category “obligatory” contexts of use or intensional subjunctive) will be the easiest to learn for L2 speakers (Perez-Cortes, 2016).

While L2 learner’s knowledge of the Spanish subjunctive has also been studied from multiple theoretical frameworks, Second Language Acquisition (SLA) and sociolinguistic/variationist research are the most widespread. The SLA literature has mostly examined production and interpretation of mood in advanced and native-like L2 learners. Despite this, studies have yielded contradicting results: while some have reported that, even after being immersed in the L2, the subjunctive remained a difficult structure for learners to use in a native-like way (Cheng & Mojica-Diaz, 2006; Bonilla, 2015; Isabelli & Nishida, 2005), a few others have provided evidence that learners can indeed acquire mood selection across the same contexts as monolingual speakers (Iverson et al., 2008). Importantly, however, L2 learners’ performance is overall better in obligatory
than in optional contexts. For instance, Massery and Fuentes (2014) tested identification and use of mood selection in deontic (i.e., obligatory contexts) as well as epistemic and epistemological predicates (i.e., optional contexts) in L2 learners (N= 150) at difference proficiency levels using a mood conjugation task. While results from obligatory subjunctive selection in deontic environments yielded high scores across groups (ranging from 73.2% in beginners to 91.6% in advanced learners) participants scored much lower in contexts allowing for mood alternations (ranging from 32.6% to 55.4%), where even advanced learners performed at chance. Finally, it is also important to mention that although some studies have compared heritage speakers’ and L2 learners’ knowledge of mood, results also diverge here. For example, both Mikulski (2006) and Montrul and Perpiñán (2011) compared heritage speakers’ and L2 learners’ use of mood in obligatory contexts using (i.e., volitional constructions). While Mikulski found that heritage speakers performance was significantly better than that of L2 learners, Montrul and Perpiñán found the opposite; differences in the results of reported in these studies could be due to the methodology employed or the participant sample recruited. Still, as explained in Chapter 1, subsection The linguistic knowledge of heritage speakers, the extant literature comparing heritage speakers to L2 learners has shown that because of the acquisitional differences between these populations, learners tend to slightly outperform heritage speakers in tasks that tap into morphosyntactic knowledge, while heritage speakers’ command of phonology is more native-like (for reviews see Montrul 2015; Polinsky, 2018; Polinsky & Scontras, 2019).

Rather than relying on prescriptive norms like the SLA studies discussed above, the variationist literature has focused on analyzing the frequency of choice of a given variant (indicative vs. subjunctive) and the linguistic predictors that modulate that choice during natural language use in learners and natives (Gudmestad, 2018). Among these
studies, the work by Aarnes Gudmestad is perhaps the most relevant. For example, Geeslin and Gudmestad (2008, 2010) used a semi-guided interview written contextualized task (WCT) to examine production of the Spanish subjunctive in 16 advanced L2 learners and 16 native speakers. Based on work by Collentine (1995) and Lubbers Quesada (1998), the authors established a series of linguistic factors that they predicted would be relevant in subjunctive use among L2 learners and natives, namely, (1) the semantic category of the governor, (2) the morphological regularity of the subordinate verb, and (3) the presence of futurity in the clause. The results indicated that even though the learners showed lower overall rates of subjunctive production than the natives, the same linguistic factors modulated mood selection production in both groups of participants; for instance, both learners and native speakers produced more subjunctive forms with volitional governors and with irregular subordinate verbs, indicating that these factors favor subjunctive selection in L2 grammars. In a 2012 study, Gudmestad applied these findings to an analysis of L2 interlanguage development with the goal of showing how learners acquire the ability to vary their use of verbal moods in Spanish. To do so, she compared the performance of L2 learners at five different proficiency levels and monolingual speakers in three oral production tasks designed specifically for the elicitation of mood use. The same linguistic factors discusses above (i.e., semantic category, morphological regularity, and futurity) were included as predictors in a multivariate analysis of participants’ responses. Regarding proficiency, it was found that learners tended to become more nativelike as their proficiency improved. However, their development of mood selection was not entirely linear; while more advanced L2 learners did use the subjunctive more frequently than beginner learners, intermediate learners showed a temporary drop in target-like use of the subjunctive (see Gudmestad,
2012, 2018 for in-depth discussion). As for the effect of the linguistic predictors examined, results indicated that all of the linguistic features helped to identify differences in mood use among the five L2 groups. Semantic category and form regularity where significant predictors at all levels of proficiency, while futurity and morphological regularity were only significant for the intermediate and advanced groups, although their effect was stronger for the advanced learners.

The results of the studies discussed so far are extremely relevant for the L2 literature because they provide empirical evidence for the time-course of acquisition of mood selection and the variables that modulate it in L2 Spanish. Just like use of subjunctive increased as proficiency level increased, so did more linguistic factors influence learners’ mood use. Thus, the order in which the linguistic features predicted learners’ use of verbal moods can be used to identify stages of development for the contexts in which they begin to make mood distinctions. As Gudmestad (2012) points out, while her studies confirm that the rate of acquisition of the subjunctive is relatively slow given that this is a late-acquired grammatical structure (Birdsong, 1992; Slabakova & Montrul, 2003), the data still support the fact that adult learners can indeed acquire near native-like knowledge of mood use for similar results (see: Gudmestad, 2006, 2013, 2014, 2015; Kanwit & Geeslin, 2014).

In all, the studies discussed so far show that variationism can produce exceptionally rich descriptions and explanations of the linguistic knowledge of L2 learners and its evolution by investigating variables at the word, sentence and discourse level. This highlights the fact scholars need to look beyond mere frequency of subjunctive use differences between L2 learners and native speakers.
Recapitulation

Characterizing heritage grammars in terms of *simplification, reduction, attrition* or *incompleteness* is prevalent in the literature, despite of increasing evidence to the contrary. In a further attempt to challenge these notions, rather than studying “younger” (Torres Cacoullos & Travis, 2018) and more heterogenous heritage communities, I propose to study the long-term bilingual community established in Albuquerque, New Mexico. In the following section of this chapter, I provide a brief but informative description of the New Mexican community of heritage speakers under study in this dissertation. This will be critical for the reader to achieve a more nuanced understanding of the participant sample examined, which will, in turn, help better frame the results reported.

**The Spanish of New Mexico and its heritage speaker community**

Spanish has been spoken in New Mexico, especially in the north of the state, for over 400 years, well before English, (Torres Cacoullos & Travis, 2018), making this state home to the oldest Spanish-speaking community in the US (Lipski, 2008). Before the state of New Mexico was established, its territory was occupied by the Spaniards for a span of 300 years (1521-1821; Espinosa, 1909) and did not become part of the US until the Treaty of Guadalupe Hidalgo was signed in 1848 (Bills & Vigil, 1999). Critically, while annexation to the US entailed a change in language policy that made English the *de facto* language of the region, both English and Spanish have continued to flourish across the state (Halberstadt, 2017; Travis & Villa, 2011). Today, New Mexico has the highest proportion of Hispanics in the US, 48% in comparison to the 17% national average (Torres Cacoullos & Travis, 2018: 16; Travis & Villa, 2011; see also United States Census Bureau 2014). New Mexico is also set apart from other states because its Hispanic population
overwhelmingly includes individuals native to the region, rather than Mexican-born immigrants. As shown in Figure 1-2 below, first-generation immigrants across the state account for only 16% of the population, in comparison to the majority of residents who are New Mexico-born (66%; García-Acevedo, 2000; Gonzales Berry & Maciel, 2000; see also United States Census Bureau, 2014).

**Figure 1-2.** Place of birth of Hispanics in New Mexico

Modern-day New Mexican Spanish is a variety spoken by the descendants of the original Spanish-speaking settlers. It was developed in what is now US territory several centuries ago and was subsequently isolated from more recent innovations of Mexican Spanish. Lipski, for instance, calls it “relatively untouched colonial Mexican Spanish” (2008: 78). As explained by Torres Cacoullos and Travis (2018), many of the individuals who settled New Mexico had been born in the *New World* and brought with them a language already evolving from Peninsular Spanish varieties (p.18). Thus, while New Mexican Spanish follows the morpho-syntactic norms of Mexico, as opposed to Spain...
(e.g., the Preterit canté, ‘I sang’, is the default past perfective rather than the Present Perfect he cantado, ‘I have sung’), it has developed its own lexical items and phonetic features (Bills & Vigil, 2008: 15; Brown, 2005; Garland Bills & Neddy Vigil, 2008). For instance, a number of archaic features can be found in New Mexican Spanish: trujo vs. trajo, ‘he brought’, or vide vs. vi, ‘I saw’ (Espinosa, 1911; Lipski, 2008). In addition, as in any bilingual community, anglicisms are also present: troca, ‘truck’, or quara/cora, ‘quarter’, (25 cents).

After the railroad arrived in New Mexico in 1878, the number of English-speaking people in the state gradually grew from under 10% in 1880, to about 50% by the 1940s (Bills & Vigil, 1999; Fernández-Gibert, 2010). Once New Mexico officially became a state and English became the official language, Spanish was in increasingly displaced, especially in schools, depriving the population of access to education in the native language of their communities (Gonzales, 1999). It has been argued that through the first half of the 20th century, public schools were an instrument for the imposition of English through suppression of Spanish (Gonzales-Berry, 2000). In their book, Torres Cacoullos and Travis (2018) provide testimonies from NMSEB participants on the experiences at school. All participants agreed that they were punished for speaking Spanish, laughed at for their English skills, and denied the opportunity to code-switch (p. 20).

Unfortunately, the repression of Spanish in the school system has worked the detriment of the local variety, which is stigmatized in comparison to the standard Spanish that is taught in schools (Bills & Vigil 2008; Gonzales-Berry, 2000). It is oftentimes the case that New Mexican Spanish is deemed “incorrect” or “uneducated” because of its archaic features (Lipski, 2008) and considered inferior even by its own speakers. On this topic, Torres Cacoullos and Travis explain that:
The disparagement of the local variety feeds linguistic insecurity [which] is discernable among NMSEB participants who regard New Mexican Spanish as not “real Spanish” [...] Thus, New Mexican Spanish is undervalued in comparison with English and monolingual Spanish, both that of the Mexican immigrant population and the textbook variety taught in schools, a stigmatization that has propelled it toward highly endangered status. (p. 25).

The fact that Spanish and English have co-existed in New Mexico for over 150 years, make this state a site of intense language contact. As many other states in the Southwest, a shift to English is underway (Bills, Hernández Chávez & Hudson 1995; Torres Cacoullos & Travis, 2018). There is evidence of intergenerational language loss as seen in the fact that 41% of US-born New Mexicans report speaking no Spanish. Jenkins (2013) reports that New Mexico has lost significant ground in language loyalty; while in the 1980 figures for several counties had language loyalty measures (i.e., maintenance) over 101%, the 2000 data show no county with a measure greater than 84% (p. 15). There are, nevertheless, Hispanic New Mexicans resisting the shift to English in the northern counties. Here, Torres Cacoullos and Travis report that approximately two thirds of the Hispanic population speak Spanish in the home, while the percentage of native Hispanics who “Speak only English” is as low as 29%-36% (2018: 25; see also United States Census Bureau 2014).

Recapitulation
While a shift to English is indeed under way, New Mexico is the home to the oldest bilingual community in the US, which makes it an ideal setting to study of mood selection in heritage grammars.
CHAPTER 2: MAXIMIZING THE EXPLANATORY POWER OF QUESTIONNAIRE DATA FOR THE STUDY OF HERITAGE POPULATIONS.

Introduction

There is now a long history of using Language History Questionnaires (LHQs) to measure the impact of bilingual experience on language acquisition and processing. However, despite the widespread use of LHQ data to explain and predict language outcomes, generalizing findings is oftentimes difficult given the lack of consistency in the criteria used to characterize and differentiate bilingual populations across studies (de Bruin, 2019; Lehtonen et al., 2018; Surrain & Luk, 2019). Metanalyses and reviews suggest that studies diverge extensively in the extent to which they report LHQ data, which can make comparing and synthesizing studies problematic (see Surrain & Luk, 2019, for a comprehensive review; also, Hulstijn, 2012, for proficiency; Birdsong, 2016; Birdsong & Vanhove, 2016, for language acquisition history and dominance; Green & Abutalebi, 2013, Paradis & Nicoladis, 2007, for contextual factors) This problem becomes even more challenging in the case of studies that focus on highly variable bilingual populations such as heritage speakers (Benmamoun, Montrul & Polinsky, 2013; Putnam & Sánchez, 2013; Rothman, Tsimpli, & Pascual y Cabo, 2016; Silva-Corvalán, 1994; Zyzik, 2016). Although the nature of the linguistic knowledge of heritage speakers has been investigated and discussed extensively, the inherent variability of this population is one of the greatest challenges scholars encounter when studying heritage grammars (Kanno, Hasegawa, Ikeda, Ito, & Long, 2008; Kondo-Brown, 2005; Valdes, 2004). The source of this variability stems from the extraordinary diversity of the bilingual experience of heritage speakers (e.g., differences in age of acquisition (AoA), quality and quantity of input, interactional
contexts of language use, literacy level in the heritage language, sociopolitical situation, cultural identification, community of practice, etc.).

Differences in the way heritage populations, and their comparison groups, are characterized can have a direct impact on the results reported in studies. For instance, Montrul (2009) and Giancaspro (2017) used similar offline tasks to investigate how heritage speakers of Spanish at different proficiency levels comprehended and produced variable subjunctive use in relative clauses. According to Montrul’s results, even at advanced levels of proficiency, heritage speakers were significantly less accurate in choosing subjunctive forms than the control group, and had significantly lower ratings for subjunctive forms. By comparison, Giancaspro found that the heritage group that he recruited behaved very similarly to the control group in how they comprehended and produced the subjunctive and maintained “systematic knowledge of the morphosyntactic features that underlie subjunctive mood morphology” (p.289).

It is possible that the discrepancy in the results of these two very similar studies might be caused, at least in part, by the way in which their heritage samples were characterized. While the proficiency of heritage speakers in both studies was measured using the Diploma del Español como Lengua Extranjera (DELE), the LHQ data provided for each sample differed considerably. On the one hand, Montrul describes the heritage speakers in her sample in terms of five factors: (1) the demographics of their parents; (2) the AoA of the majority language (i.e., English); (3) their literacy in the majority language; and (4) their use of the heritage language at home, and (5) their dominance. On the other hand, for his sample, Giancaspro reports (1) parents’ demographics, generation; (2) AoA

---

6 Montrul (2009): elicited production task, written morphology recognition task and sentence conjunction judgment task; Giancaspro (2017): contextualized elicited production task, contextualized acceptability task and mood preference task.
of the majority and heritage language, (3) use of and exposure to the heritage language with family members and friends, and (5) literacy both in the majority and heritage language. In addition, while Montrul only states that the control group was comprised of “a total of 23 fluent native speakers of Spanish” (p.249), Giancaspro provides the same information about his control group, who were “Spanish-dominant bilinguals who began acquiring English at age 13 or later” (p. 103).

The fact that less quantitative and qualitative information about participants is provided in the study by Montrul (2009) makes it difficult to assess why its findings diverge from those of Giancaspro (2017). For instance, it is possible that the heritage speakers in Montrul’s sample had lower literacy in the heritage language in comparison to Giancaspro’s sample. Given that the subjunctive is more frequent in written formal registers, this might have impacted their performance. Likewise, Montrul’s study provides insufficient information about participants’ use of and exposure to the heritage language; it is possible that, in Montrul’s sample, participants’ overall use of the heritage language was lower, which would have impacted their knowledge of the subjunctive. Heritage speakers who use the heritage language more frequently, with more interlocutors and in more contexts have exposure to more extensive input which is especially necessary infrequent constructions such as the subjunctive, as more input strengthens the representation of this construction in the heritage grammar.

This comparison highlights the heterogeneity of participants under the same “heritage speaker” label and the consequences that this has on the interpretation and extrapolation of results across studies. Results from studies on heritage grammars need to be understood not just in terms of proficiency and AoA, but also in the context of individuals’ socially-situated experience with language use. This is especially relevant given that there is strong evidence that psychosocial variables and biographical factors
impact heritage grammar outcomes (see: Ducar, 2012, for motivation; Allard & Landry, 1992, for language attitudes; Rivera-Mills, 2012, for generation; Swender, Martin, Rivera-Martínez & Kagan, 2014, for years of formal education in the heritage language). For instance, Torres, Estremera & Mohamed (2019) used a spoken elicited imitation task and a written untimed grammaticality judgment task to test heritage speakers of Spanish knowledge of a series morphosyntactic structures that have been shown to be problematic in heritage acquisition (i.e., noun and adjective gender agreement, case marking with dative psych verbs, animate direct objects, and double object clitics, articles with generics, aspect and mood). Participants also completed 4 questionnaires to report a series of common psychosocial and biographical factors in relation to the heritage language (e.g., AoA, generation, motivation, willingness to communicate, language attitudes and frequency of use). The results of mixed-effects regression models revealed that variables such as sequential bilingual status, willingness to communicate, generation and motivation contributed significantly to participants’ performance on both tasks such that participants with higher motivation and who belonged to an earlier generation had higher scores.

In all, these findings underscore the explanatory power of LHQ data and the importance of providing fine-grained descriptions of participant samples in order to make findings more interpretable and generalizable. Still, LHQs have their own drawbacks: a common problem faced by scholars using LHQ data is that questionnaires often produce a number of highly correlated variables, which leads researchers to “cherry pick” (de Bruin, 2019) variables based on their own judgement. I propose that what is needed is a method that systematically disentangles LHQ variables to help understand what questions are more informative for the description and differentiation of bilingual populations. Here, I use Principal Component Analysis (Flury, 1988; Hotelling, 1933;
Jackson, 1991; Jolliffe, 2002; Jolliffe & Cadima, 2016) as a method to empirically quantify dimensions of bilingual experience in order to maximize the explanatory power of LHQ data. The goal is to provide an evidence-based characterization of heritage speakers by examining which language-experience factors best explain the variability inherent to this population. In the following section, we review the most common dimensions of bilingual experience included in questionnaires and discuss how heritage speakers vary along these dimensions.

Dimensions of bilingual experience and heritage speaker variability

While a considerable number of LHQs have been created over the years, these questionnaires tend to include similar variables to characterize bilinguals’ linguistic experience along core dimensions. An examination of several of commonly used LHQs (Beaudrie & Ducar, 2005; Ikizer & Ramirez-Esparza, 2017; Gollan, Starr & Ferreira, 2015; Kennedy, 2012; MacIntyre, Baker, Clément, and Conrod, 2001; Marian, Blumenfeld & Kaushanskaya, 2007; Lev-Ari & Shao, 2017; Li, Zhang, Tsai & Puls, 2020; Luk & Bialystok, 2013; Rodriguez-Fornells, Krämer, Lorenzo-Seva, Julia Festman & Münte, 2012; Torres, Estremera & Mohamed, 2019) reveals that the most frequent dimensions and the variables included are:

(1) *Language background dimension*: factors related to age of acquisition (AoA), manner of acquisition, proficiency, and dominance.

(2) *Language use dimension*: factors having to do with contexts of use and interactional network.

(3) *Affective dimension*: affective factors such as motivation, attitude identification.
Variables related to the language background dimension are perhaps the most studied within bilingual research. Heritage speakers vary in the AoA of the non-heritage language which, in turn, modulates proficiency and dominance in the heritage language. For instance, some heritage speakers are child immigrants; oftentimes, this causes the heritage language to become the minority language in the country they emigrate to, which drastically reduces input in this language, while the majority, non-heritage language input increases (see Hoff & Core, 2013, for a review on input and language development in heritage speaker children). It has been shown that early exposure to the majority language oftentimes leads to a shift in dominance from the heritage to the non-heritage language (Basnight-Brown & Altarriba, 2007; Harris, Berko Gleason, & Aycicegi, 2006; Heredia & Altarriba, 2001; Flores, Santos, Jesus & Marques, 2017; Kohnert, Hernández, & Bates, 1998; Martin, Altarriba & Kazanas, 2020; Montrul & Foote, 2012; Montrul & Ionin, 2012). Other heritage speakers are second generation bilinguals and are exposed to both languages in a more balanced fashion from birth, until they begin to receive formal education in the majority language. Overall, evidence seems to indicate that the earlier heritage speakers are exposed to the non-heritage language, the less proficient they tend to be in the heritage language (Au et al., 2002; Montrul, 2002, 2009; Montrul et al., 2008).

Proficiency and dominance are also modulated by manner of acquisition of the heritage language, which indexes whether a heritage speaker acquired the heritage and non-heritage languages in naturalistic (e.g., home) or formal environments (e.g., school; see Polinsky & Kagan, 2007, for a review). Thus, heritage speakers vary in their literacy in the heritage language. Most heritage speakers acquire the heritage language primarily orally in naturalistic settings and usually only receive formal education in the non-heritage language. However, other heritage speakers start taking classes in the heritage
language once they start college, and still others receive formal education both in the heritage and non-heritage language their whole lives. This highlights the fact that there is considerable variability in the manner of acquisition of both heritage language and non-heritage language (Kondo-Brown 2009; Kupisch, 2013; Montrul, 2013; Montrul et al., 2008; Pires & Rothman 2009; Rothman 2007).

Finally, it is important to mention that, in LHQs, proficiency is commonly assessed by means of self-ratings. However, there is evidence that self-ratings of bilingual proficiency can be misleading; factors such as language combination, cultural identification, sociopolitical status, language dominance and affective factors such as anxiety can impact how bilinguals self-rate their proficiency (MacIntyre, Noels & Clément, 1997; Tomoschuk, Ferreira & Gollan, 2019). This makes self-ratings especially problematic to use with heritage populations, which are oftentimes the target of a double stigma: not only are they speakers of a minority language in the society they live in but, in many cases, they have learned varieties of the heritage language that are stigmatized in the home country (Valdés, 1997; Poplack & Torres Cacoullos, 2015). Thus, it is common to observe that heritage speakers rate their linguistic skills overall lower than other bilingual populations would, which makes their self-ratings highly unreliable (Giancaspro, 2017; Perez-Cortes, 2016; Perez-Cortes, Puntam & Sanchez, 2019).

The language use dimension of LHQs encompasses the frequency and manner in which bilinguals use their languages, which is reflected in their contexts of use and their interactional networks. As regards heritage speakers, frequency of use of the heritage language has been identified as crucial for increasing and maintaining proficiency (Bedore et al., 2012; Gathercole & Thomas, 2009; Kenji & D’Andrea, 1992; Tsai, Park, Liu & Lau, 2012). The way the heritage language is used in different settings and activities varies to a large extent based on heritage speakers’ community of practice. Heritage
speakers whose families live in largely monolingual communities have less chances of being exposed to the heritage language outside their home and, thus, use it less overall, which can negatively impact their proficiency. By comparison, heritage speakers who belong to thriving bilingual communities have the chance to use the heritage language with a wide variety of people in a broad range of contexts (Aalberse & Muysken, 2013; Valdés, 2001). For instance, Carreira (2004) exemplifies that the conditions that characterize heritage communities in a city like Miami, which has a high density of Latinos, are radically different from those that apply to communities in a place like Birmingham Alabama, which has a small and more spread-out Latino population. Evidence also suggests that heritage speakers with larger and more heterogenous interactional networks are more proficient in the heritage language and have more balanced dominance in both languages (Gathercole & Thomas, 2009; Kenji & D’Andrea, 1992; Hulsen, de Bot, & Weltens, 2002; Place & Hoff, 2011; Rivera-Mills, 2012). In fact, the same pattern has been found for heritage speakers who use the heritage language in more diverse settings, even if they interact with less interlocutors (Gullifer & Titone, 2019).

Finally, variables within the affective dimension (e.g., motivation, attitude and identification) also have a significant impact on bilingual grammars but are not so often reported in the literature. Once again, heritage speakers vary along this dimension depending to a large extent on the community they belong to. For instance, home and community environments that value maintenance of the heritage language can promote children’s identification with the culture of the heritage language and their motivation to learn it. In turn, more positive attitudes, higher motivation, and stronger identification with the heritage language led to overall higher proficiency (Beckstead & Toribio, 2003; Beudrie & Ducar, 2005; Carreira, 2003; Carreira & Kagan, 2011; He, 2006; MacIntyre, Baker, Clément & Donovan, 2002). Affective factors also impact heritage language
maintenance: heritage speakers in communities with more positive language attitudes have a higher tendency to maintain the heritage language across more generations. For instance, studies conducted within long-established bilingual communities, such as the one in Albuquerque, New Mexico, have shown that the heritage language can remain vital even after five generations (Torres Cacoullos et al., 2018; LaCasse, 2018).

In all, the literature reviewed so far showcases the inherent variability of heritage populations across a series of core dimensions of bilingual experience. In the next section, I describe the LHQ used in the present study and explain how it was designed to create the linguistic profile of participants.

**Method**

**Participants**

A group of 33 New Mexican heritage speakers of Spanish participated in this study (27 females, mean age 20.13 years old, SD= 1.67, range= 5). All participants were in the process of completing a bachelor’s degree at the University of New Mexico at the time of data collection; participants were recruited from a range of upper-level courses. Besides the sociolinguistic questionnaire, participants completed three behavioral tasks that were proxies for language proficiency: Verbal Fluency task, Bilingual Picture Naming task and Elicited Imitation Task. In the following subsections, a concise summary of participants’ demographics, language background and language use is provided to acquaint the reader with the heritage sample at hand. In addition, each of the behavioral tasks is described to provide a nuanced picture of participants’ proficiency.
Demographic and linguistic characteristics

Given the high interspeaker variability inherent to heritage populations, it is particularly important to not only provide information about language background and use, but also a detailed demographic characterization of the speakers in the sample and the community they belong to.

The demographic questions revealed that 26 out of the 33 participants were born and raised in Albuquerque, New Mexico, while the remaining seven were born in Mexico and had moved to the US as infants (mean age of immigration 2.50 years old; SD= 1.38, range= 4). Participants’ parents were born mostly in New Mexico (N= 50/66); the remaining 16 were born in Mexico and had been living in Albuquerque an average of 25.44 years (SD= 5.98, range= 22.5). Similarly, most of the participants’ grandparents were born in New Mexico (N= 52/66), while the rest had been born in Mexico; of those, 10 had lived in the US for an average of 25 years (SD= 5.10, range= 20), while the rest still resided in Mexico. This information underscores that the majority of the sample was fairly homogenous in terms family demographics and were established members of the New Mexican bilingual community in Albuquerque (see Chapter 1, subsection The Spanish of New Mexico and its heritage speaker community, for an in-depth description of this community). Regarding participants’ parents’ socioeconomic background, Table 2-1 below shows the distribution of education levels across the sample:

<table>
<thead>
<tr>
<th>Table 2-1. Education level of participants parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary school</td>
</tr>
<tr>
<td>Middle school</td>
</tr>
<tr>
<td>High school</td>
</tr>
<tr>
<td>Bachelor’s</td>
</tr>
<tr>
<td>Master’s</td>
</tr>
<tr>
<td>Ph.D.</td>
</tr>
</tbody>
</table>
The questions eliciting linguistic information were designed to characterize participants’ life-long linguistic experience with Spanish and English; information was elicited about participants’ age of exposure to Spanish and English, perceived dominance in both languages, and language use in different contexts. In addition, participants were asked to use a 7-point scale (from 1-Very low to 7-Perfect) to self-rate their proficiency in reading, writing, speaking and understanding each language; the self-rating was administered twice during the data collection procedure: first at the end of the first session, and then at the end of the second session. This was done to roughly ascertain the relative reliability of the measure.

The questions collecting language use data showed that even though all participants reported having been exposed to Spanish significantly earlier than English \((t(40) = -7.10, p < .001)\), slightly over half of the participants considered themselves English to be their dominant language \((N = 19/33)\). It follows that self-ratings revealed that participants felt more proficient in English than in Spanish on average (Session 1: \(t(60) = -2.06, p = .04\); Session 2: \(t(55) = -2.21, p = 0.03\)). The self-ratings did not differ significantly between the two sessions (for Spanish \((t(50) = -1.65, p = 0.10)\); for English \((t(58) = -1.02, p = .31)\). Finally, participants reported using Spanish and English in their daily lives almost to the same degree \((t(62) = 1.88, p = .06)\). Participant characteristics are summarized in Tables 2-2 and 2-3.

| Table 2-2. Participants’ self-reported characteristics |
|-------------------------------|---|---|
| Measure                       | Mean | SD  |
| Age (years)                   | 20.13 | 1.67 |
| Age of acquisition: Spanish   | 0.26  | 2.38 |
| Age of acquisition: English   | 3.61  | 2.38 |
| Spanish Use                   | 46.38% | 15.42 |
| English Use                   | 53.63% | 15.40 |
Participants were also asked to specify which language they used most frequently: at home, school, work, free time and when engaged in hobbies (they used a 7-point Likert scale from 1-Never to 7-Always); the purpose of this question was to gain a measure of participants’ diversity of language use across different social contexts. As can be seen in Table 2-4 below, use of English is higher across all contexts except for Home, as heritage speakers tend to use Spanish mostly with their family members (Gollan, Starr & Ferreira, 2009). Use of Spanish is lowest at Work given that English is the majority language and, therefore, used more frequently in formal and professional contexts.

### Table 2-4. Diversity of language use across social contexts

<table>
<thead>
<tr>
<th>Context</th>
<th>Spanish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Home (/7)</td>
<td>6.78</td>
<td>0.71</td>
</tr>
<tr>
<td>Work (/7)</td>
<td>3.66</td>
<td>1.75</td>
</tr>
<tr>
<td>School (/7)</td>
<td>5.32</td>
<td>2.18</td>
</tr>
<tr>
<td>Free time (/7)</td>
<td>4.44</td>
<td>2.02</td>
</tr>
<tr>
<td>Hobbies (/7)</td>
<td>4.19</td>
<td>1.87</td>
</tr>
</tbody>
</table>
Behavioral tasks

Accurately assessing language proficiency is challenging in the case of heritage speakers. Traditional measures such as standardized tests of proficiency (e.g., Diploma de Español como Lengua Extranjera, Bilinguals English Spanish Assessment) are not truly reflective of their proficiency level. Heritage speakers’ literacy in the heritage language is oftentimes reduced which entails that their metalinguistic skills and knowledge of standardized forms of the heritage language are limited. Given that standardized tests have not yet been normed with heritage populations, they intrinsically put heritage speakers at a disadvantage.

To address this problem, the present study used three implicit, oral behavioral tasks as proxies for proficiency in both Spanish and English: Verbal Fluency, Bilingual Picture Naming and Elicited Imitation. Given that heritage speakers’ experience with the heritage language is mostly informal and oral, these tasks are more appropriate to measure proficiency in heritage populations, as they are more ecologically valid in relation to speakers’ experience with the heritage language.

Verbal fluency task. Verbal fluency is a task employed to evaluate verbal ability including lexical knowledge (Weckerly, Wulfeck, Reilly, 2001; Federmeier, Kutas & Schul, 2010). It is used in this study as a proxy for proficiency because it is expected that participants with larger vocabularies will be more proficient in the language than participants with smaller vocabularies. In this task participants are given a category (e.g., FRUIT) and are asked to name as many exemplars as possible (APPLE, ORANGE, BANANA) in a predetermined amount of time. Participants in this experiment were asked to name eight categories (body parts, colors, furniture, fruits, animals, vegetables, clothing, and musical instruments). Thirty second were allotted to name exemplars in
each category. Presentation of the categories was counterbalanced by language so that if one participant named exemplars of FRUIT in Spanish, the next participant would name exemplars of the same category in English. In practical terms, this means that participants named four categories in one language and the remaining four in the other language. The total number of words participants named in each category counted cumulatively towards a final score. Words that were unintelligible or repeated were excluded. In addition, words that were derivations of another (e.g., blue: light blue, dark blue, baby blue, navy blue) were also excluded. Results showed that participants named significantly more exemplars in English (M= 41.75) than in Spanish (M= 37.75; t(63)= 2.27, p = 0.02).

**Picture naming task.** The picture naming task is used to evaluate lexical knowledge and lexical access ability (Gollan, Montoya, Fennema-Notestine & Morris, 2005; Lee & Williams, 2001). As in the verbal fluency task, it is used here as a proxy for proficiency because it is expected that participants with larger vocabularies will be more proficient in the language than participants with smaller vocabularies. In most versions of a picture naming task, participants are shown pictures on a computer screen one by one and are asked to name each as quickly and accurately as possible. In this experiment, we adapted a version of the picture naming task used by Gollan, Montoya, Cera & Sandoval (2008). Participants were asked to name 90 black and white line drawings. Half of the pictures were presented in a Spanish block and the other half were presented in an English block. The order of the language was counterbalanced by participant such that if one participant completed the first block in English and then in Spanish, the next participant completed the task in the reverse order. Items were matched for frequency and concreteness across languages, as shown in Table 2-5 below.
Table 2-5. Characteristics of picture naming items

<table>
<thead>
<tr>
<th></th>
<th>Spanish</th>
<th></th>
<th>English</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Range</td>
<td>M</td>
</tr>
<tr>
<td>Log Freq. (per mill.)</td>
<td>1.06</td>
<td>.58</td>
<td>.7-2.61</td>
<td>1.06</td>
</tr>
<tr>
<td>Concreteness</td>
<td>6.19</td>
<td>0.44</td>
<td>5.04-7.00</td>
<td>5.81</td>
</tr>
</tbody>
</table>

English lexical frequency and concreteness norms were derived from the English Lexicon Project (Balota et al., 2007). The lexical frequencies and concreteness values of the Spanish names were obtained from the EsPal database (Duchon, Perea, Sebastián-Gallés, Martí & Carreiras, 2013). Reaction times (RTs) were recorded, and responses were coded for accuracy. A response was considered accurate if it matched the intended target name. RTs that were associated with inaccurate responses and registration errors (e.g., hesitations and repetitions) were excluded. Results showed that, overall, participants were significantly more accurate in English (M= 92.08%) than in Spanish (M= 82.10%; t(40) = 4.75, p < 0.01), and significantly faster to respond in English (M= 1310.09 ms) than in Spanish (M= 1053.45 ms; t(50) = -2.14, p = 0.02).

**Elicited imitation task.** The purpose of this task is to measure implicit knowledge and aural/oral proficiency (Bowden, 2016). In this task, based on the version by Solon, Park, Henderson & Dehghan-Chaleshtori (2019), participants listened to 36 Spanish sentences one at a time and were asked to repeat them after a tone. The pause after each sentence before repetition was set at 2 s, and the tone lasted 0.5 s, for a total of 2.5 s between the end of each sentence and the window for the start of the repetition, which lasted 9 s. Crucially, as the tasks progressed the sentences increased in number of syllables from 7 to 27, thus making recall progressively more difficult. Table 2-6 below shows the total number of sentences for each syllable range:
Most sentences were subordinate (N= 25), but there were also simple (N = 6) and coordinate sentences (N = 5). For instance, a simple, 7-syllable sentence was *Quiero cortarme el pelo* (‘I want to cut my hair’), while a subordinate, 27-syllable sentence was *Acabamos de volver del supermercado donde las ofertas eran muy interesantes* (‘We have just come back from the supermarket were the offers were very interesting’).

The task was scored in two ways: first, the repetition of sentences was scored in terms of lexical and structural accuracy; second, repetition was scored for accentedness, fluency and comprehensibility. The first type of scoring was conducted both by the experimenter and two individuals who were blind to the purpose of the task. The inter-rater reliability rate of the scorers was 86.84% (Isaacs & Thomson, 2013), indicating that there was high similarity in their ratings. A scale from 0-4 based on how faithful participants’ repetition was to the meaning and content of the original sentence was used (Solon et al., 2019):

- 0 = Silence, unintelligible content, or only one content word.
- 1 = Half or less of content repeated; important content is left out; meaning may be unrelated or opposed to stimulus.
- 2 = More than half of content preserved; slight changes in content that make content inexact, incomplete, or ambiguous.
- 3 = Meaning preserved; use of synonyms or changes in grammar that do not affect meaning.

**Table 2-6. Syllable distribution in EIT**

<table>
<thead>
<tr>
<th>Syllables</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>23</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>36</td>
</tr>
</tbody>
</table>


4 = Perfect repetition.

Given that there were 36 sentences, each of which could get a maximum score of 4, the highest overall score possible in the task was 144. Results showed participants scored 139.40 on average (SD= 4.43, range= 126-144).

The second type of scoring was only conducted by the same two same individuals who scored the task the first time, as both native speakers of Mexican Spanish. The inter-rater reliability rate of the scorers was 83.55%, once again indicating high similarity in their ratings. In this case, participants’ repetition was scored using 5-point scales of the following parameters (based on Kang, 2010):

- **Accentedness**: how different does the speaker sound from a native speaker born and raised in Mexico. A 5-point scale was used ranging from 1-Heavily accented to 5-Not accented at all.
- **Fluency**: how smooth the speaker’s oral delivery is based on pausing, hesitations, fillers, or speech rate. A 5-point scale was used ranging from 1-Very dysfluent accented to 5-Very fluent.
- **Comprehensibility**: how easy to understand is the speaker. A 5-point scale was used ranging from 1-Very hard to understand to 5-Very easy to understand.

Results showed that participants’ mean accentedness was 4.50 (SD= .63, range= 2.9-5), their mean fluency was 4.11 (SD= 0.44, range= 2.8-4.8) and their mean comprehensibility was 4.27 (SD= .42, range= 3.1-4.9). In summary, participants scored fairly high in this task across all measures, indicating a high level of oral and aural proficiency as well as a good command of Spanish morphosyntax.
Materials and Design

The sociolinguistic questionnaire created for this study was designed by examining a series of widely-use questionnaires whose internal validity and reliability have already been thoroughly tested\(^7\): Beaudrie and Ducar (2005), Ikizer and Ramirez-Esparza (2017), Kennedy (2012), Li et al., (2020), and Marian et al., (2007). The questionnaire created for this study consisted of a total of 70 questions; however, I will only focus on those relevant to the analyses carried out. Based on the literature discussed above, there were 57 relevant questions divided into four dimensions:

1. *Heritage Language Background* – 25 questions.
2. *Contexts of Use* – 9 questions.
3. *Interactional Network* – 7 questions
4. *Attitude, Motivation and Identification* – 16 questions

The goal of this sociolinguistic questionnaire was to build sociolinguistic profiles that accurately characterized the heritage language group in this study by systematically operationalizing their experience with language in different core dimensions of linguistic experience. The flowchart provided in Figure 2-1 below shows a breakdown of the different variables that were analyzed within each dimension (4 dimensions, and a total of 24 variables):

---

\(^7\) The validity and reliability of these LHQs have been tested by many previous studies that correlated their results with behavioral and cognitive outcomes of bilingual experience; see, for instance, Li et al., (2020:5).
Dimension 1: Heritage Language Background

This dimension was partially based on the Language Experience and Proficiency Questionnaire (Marian et al., 2007) and the Language History Questionnaire (Li et al., 2020). The goal of this dimension was to elicit detailed information about participants’ heritage language background; to do so, it included the following subsections:
a. **Demographics.** Participants were asked whether they, their parents and their grandparents were born in the US. A negative response prompted them to state the place of birth of each member and the years of residence in the US. In addition, participants also reported whether their parents and grandparents spoke Spanish, English or both.

b. **Age of acquisition and Exposure.** Participants provided the age at which they started acquiring Spanish and English, as well as the age at which they became fluent and started reading in each language. They were also asked to describe their use and exposure to both languages before and after attending school. Participants indicated which languages were used at home (Spanish, English, or both) and used a percentage range slider to gauge how much exposure to each language they had (see Figure 2-2). Whenever a percentage slide was used in the questionnaire, the sum of the percentages for all options was required to add up to 100%.

![Figure 2-2. Sample of question using percentage range slider](image)

| Before you started school, approximately what percentage of the time were English and Spanish used at home? The sum of the percentage for both language has to add up to 100%. |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
| English | | | | | | | | | | |
| Spanish | | | | | | | | | | |

c. **Proficiency and Dominance.** Participants were asked to report their dominant language (the strongest language; see Treffers-Daller, 2011), and to provide self-
ratings of proficiency in reading, writing, understanding, and speaking using a 7-point scale (1= “Very low” , 7= “Perfect”).

**Dimension 2: Contexts of Use**

This dimension was partially based on a questionnaire by Beaudrie and Ducar (2005) and elicited information on participants’ linguistic activity in five social contexts (Home, Work, School, Free time and Hobbies). Unlike past questionnaires, we were interested not just in how often participants used their languages in each context, but also in how many interlocutors they used them with. The time frame provided for all questions was weekly use, and the number of interlocutors to be reported was limited to a maximum of 10. This was done to provide participants with simple measures that would make their answers more reliable. It has been shown that questions that are easy for people to comprehend and answer maximize respondent motivation to answer carefully and thoughtfully (Krosnick, Presser, Fealing, Rugles & Vannette, 2015). The following subsections were included:

a. **Number of interlocutors per context.** Participants reported the number of interlocutors per week they spoke Spanish or English within each context (Figure 2-3).

**Figure 2-3.** Sample of question reporting number of interlocutors

<table>
<thead>
<tr>
<th></th>
<th>Home</th>
<th>School</th>
<th>Work</th>
<th>Free Time</th>
<th>Hobbies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
<td>![ ]</td>
</tr>
</tbody>
</table>

Using a scale from 1-10, estimate with how many people in total you use Spanish within each of the contexts below per week:
c. *Frequency of use in each context.* Participants used a 7-point scale to indicate how often they used each language in each context per week (Figure 2-4; 1=“Never”, 7=“Always”).

![Sample of question reporting frequency of use per context](image)

**Figure 2-4. Sample of question reporting frequency of use per context**

<table>
<thead>
<tr>
<th></th>
<th>1-Never</th>
<th>2-Rarely</th>
<th>3-Sometimes</th>
<th>4-Regularly</th>
<th>5-Often</th>
<th>6-Usually</th>
<th>7-Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hobbies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

d. *Frequency of activities.* For each language, participants used the same 7-point scale as above to indicate how often they did a series of activities in each context per week: *watching TV, listening to music, reading,* and *writing.* A global *activity* measure was derived for each context by calculating the mean of all activities.

As mentioned before, we were interested in not just *how often* participants used their languages in each context, but also in *how many interlocutors* they used them with. The *Number of interlocutors per context* and *Frequency of use in each context* data were used to derive a composite score that would be reflective of more complex patterns of linguistic experience. This composite score quantified the relative *Richness* of the contexts of use: richer contexts would be those where participants spoke a particular language more often *and* with a higher number of interlocutors. To calculate this score, the number of
interlocutors in a given context (1-10 scale) was added to the frequency of use in that same context (1-7 scale), and was then divided by the sum of both scales. This provided a proportional measure of the richness of each context per participant, the maximum of which would be 1. If a participant’s frequency of use of Spanish at Home were 7, and the number of interlocutors she used this language with were 8, then the Richness of the Home Context for this participant would be 0.88, as shown in the formula: \[ \text{participant: } \frac{(7 + 8)}{17} = 0.88 \]. Thus, individuals who used a language frequently and with a high number of interlocutors in a certain context would have a richness score close to 1. The closer the score was to 1, the higher the richness of the context of use.

**Dimension 3: Interactional Network**

This dimension was partially based on two questionnaires by Ikizer and Ramirez-Esparza (2017) and Kennedy (2012). This dimension elicited information about participants’ interactions with four different groups of interlocutors within their interactional network (Immediate family, Extended family, Friends and Acquaintances). The scales used were the same as in the previous dimension. The following subsections were included:

a. **Number of interlocutors per network group.** Participants reported the number of interlocutors per week they spoke Spanish within each network group (Figure 2-5).
b. **Number of interlocutors per activity.** Participants indicated the number of interlocutors within each network whom they interacted with on a weekly basis in Spanish when engaged in certain activities (i.e., calling on the phone, texting and getting together).

c. **Bilingualism of network.** Participants reported the percentage of interlocutors in each network who spoke Spanish, English or both (Figure 2-6). The average bilingualism of each network group was measured by calculating the mean of interlocutors who spoke both languages.
The Number of interlocutors per network group and Number of interlocutors per activity data collected were used to derive a composite score to quantify the relative Productivity of each network; that is, not just how many interlocutors per network group a participant has, but with how many of those does the participant actively engage in all three activities weekly (i.e., calling on the phone, texting and getting together). More productive networks would be those where participants spoke Spanish with more interlocutors by engaging in more activities. To calculate this score, the number of interlocutors across all three activities for a participant was divided between the maximum possible number of interlocutors (maximum of 10 interlocutors per activity = 30). This provided a proportional measure of the productivity of each network per participant, the maximum of which would be 1. If a participant phoned 10 people in her immediate family, texted 5 and got together with 10, then the Productivity of the Immediate Family Network for participant02 would be 0.83, as shown in the formula: \[ \text{participant: } \frac{10 + 5 + 10}{30} = 0.83. \] Therefore, individuals who engaged with many interlocutors in each of the three activities would

Figure 2-6. Sample of question reporting bilingualism per network group

<table>
<thead>
<tr>
<th>What percentage of your immediate family speaks only English, only Spanish or both? The sum of the percentage for both language has to add up to 100%.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
</tbody>
</table>
|------------------------------|-----------------
| English                     | English        |
| Spanish                     | Spanish        |
| Both                        | Both           |
have a high productivity score close to 1. The closer the score was to 1, the higher the productivity of the network.

**Attitude, Motivation and Identification**

This dimension was also partially based on the questionnaire by Beaudrie and Ducar (2005); it includes questions whose goal was to gauge participants’ connection with the heritage language to assess if it was overall positive or negative. Once again, because we were interested in affective factors in relation to the heritage language, questions were asked mainly about Spanish. The following subsections were included:

a. **Attitude.** Participants chose between a series of positive or negative adjectives to describe their attitude towards the heritage language (Figure 2-7). They were also given the opportunity to write freely about what they perceived to be the advantages and disadvantages of speaking the heritage language. Finally, they provided a subjective opinion on the quality of their own command of the heritage language.

![Figure 2-7. Sample of question on attitude towards the heritage language](image)

b. **Motivation.** Participants used a 7-point scale to rate their motivation to learn the heritage language (1 = “Not motivated at all,” 7 = “Extremely motivated”). They
also reported if they had ever taken courses in the heritage language and for how long.

c. Identification. Participants used a 7-point scale (1 = “Very Low”, 7 = “Perfect”) to rate their identification with Spanish and English across several cultural aspects: way of life, food, music, art, sports, and countries. Their overall identification was calculated by averaging scores across all aspects for each language (Figure 2-8).

**Figure 2-8.** Sample of question for identification

| Do you identify more strongly with English or Spanish? Use the scale below to rate the strength of your connection in the following categories for each language: |
|---|---|---|---|---|---|---|
| 1-Very Low | 2–Low | 3–Adequate | 4–Good | 5–Very Good | 6–Excellent | 7–Perfect |
| Way of life | Food | Music | Art | Sports | Countries |
| English | | | | | |
| Spanish | | | | | |

**Procedure**

Participants completed the sociolinguistic questionnaire at home through a web-based interface using Qualtrics (Qualtrics, Provo, UT). This platform allowed to present the questionnaire in a dynamic fashion that made it easier for participants to answer a high number of questions. It also made it possible to easily present questions to be answered in different ways (e.g., Likert scales, free text entry or percentage range sliders). In accordance with the regulations set by The Institutional Review Board on Human Subject
Research (IRB), participants were not forced to answer any of the questions, that is, they could continue filling in the questionnaire even if they skipped a question. For this reason, they were encouraged to take their time and do their best to answer all questions to the best of their capacity.

**Data analysis: What is Principal Component Analysis**

PCA is form of multivariate analysis that allows to reduce the complexity of high-dimensional quantitative data by identifying and summarizing the trends present in it (Flurry, 1988; Hotelling, 1933; Jackson, 1991; Jolliffe, 2002; for a recent, comprehensive review see Jolliffe & Cadima, 2016). A common problem with data sets such as LHQs that contain multiple explanatory variables is that these tend to be highly correlated. In addition to this problem, visualizing and interpreting the relationship among more than three variables in multi-dimensional space is quite challenging.

As a descriptive empirical method, PCA increases the interpretability of multivariate data with minimal information loss. Because it is descriptive, rather than inferential, this exploratory method needs not meet any distributional assumption and can be used with multiple types of numerical data (Jolliffe & Cadima, 2016). PCA reduces the complexity of high-dimensional data by “geometrically projecting them onto lower dimensions called *principal components*” (PCs; Lever, Krzywinski & Altman, 2017: 641), which correspond to the principal directions along which variation in the data is maximal. Simply put, PCs are linear combinations of the original variables that summarize latent features (i.e., underlying variation) in the data *without* reference to the researcher’s prior knowledge. While the PCA selection process maximizes the correlation ($r^2$) between the data and their projection, it produces PCs that are uncorrelated to each other, thus
maximizing the variance ($\sigma^2$) explained; Figure 2-9 below illustrates how components are geometrically orthogonal (i.e., perpendicular and, therefore, uncorrelated).

**Figure 2-9.** Principal component projection

![Principal component projection](image)

Adapted from Lever et al., (2017: 641).

PCA provides scholars with an empirical method to identify underlying informational relationships in the data. By generating linear combinations of variables that show trends of variation, PCA can substantially contribute to the recognition of relevant patterns inherent to the sample (Peres-Neto, Jackson & Somers, 2004). While the components extracted from PCA can be excellent empirical descriptors of the participant sample analyzed, it is also important to test their validity and explanatory power in relation to linguistic performance. To address this in the present study, the first PC of each dimension was correlated with heritage speakers’ performance in the behavioral task Bilingual Picture Naming (BPN; see the Participants subsection).

**Data Preparation**

Data within each dimension were analyzed in R Core Team (2020) using the packages *FactoMineR* (v. 2.23) and *factoextra* (v. 1.07) to carry out Principal Component Analysis (PCA).
In the current study, the first step in conducting PCA was to transform all categorical variables into *quantitative factors*. Dummy coding was used to create quantitative factors that represented each level of the categorical variables with a number, starting at reference level 0 (Berry, Mielker & Iyer, 1998; Daly, Dekker & Hess, 2016). Table 2-7 below shows all variables that were “dummy coded”; for instance, the first row shows that participant’s attitude towards the heritage language had *three* levels: the reference level, “Neutral”, is represented by level zero of the dummy variable, while “Positive” and “Very positive” correspond to levels 1 and 2 respectively.

**Table 2-7. Dummy coding of qualitative variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Dummy Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude towards the heritage language</td>
<td>Neutral = 0</td>
</tr>
<tr>
<td></td>
<td>Positive = 1</td>
</tr>
<tr>
<td></td>
<td>Very positive = 2</td>
</tr>
<tr>
<td>Is speaking the heritage language an advantage or a disadvantage?</td>
<td>Advantage = 0</td>
</tr>
<tr>
<td></td>
<td>Disadvantage = 1</td>
</tr>
<tr>
<td></td>
<td>Both = 2</td>
</tr>
<tr>
<td>What is your opinion of your own Spanish?</td>
<td>Neutral = 0</td>
</tr>
<tr>
<td></td>
<td>Improve = 1</td>
</tr>
<tr>
<td></td>
<td>Confident = 2</td>
</tr>
<tr>
<td></td>
<td>Bad = 3</td>
</tr>
<tr>
<td>Were you born in the US?</td>
<td>Mexico = 0</td>
</tr>
<tr>
<td></td>
<td>US = 1</td>
</tr>
<tr>
<td>What is your dominant language?</td>
<td>Spanish = 0</td>
</tr>
<tr>
<td></td>
<td>English = 1</td>
</tr>
<tr>
<td>What language were you exposed to the most before starting school?</td>
<td>Spanish = 0</td>
</tr>
<tr>
<td></td>
<td>English = 1</td>
</tr>
<tr>
<td></td>
<td>Both equally = 2</td>
</tr>
</tbody>
</table>

---

8 Data sets that *only* contain qualitative variables can be similarly analyzed using *Multiple Correspondence Analysis* (Di Franco, 2016). If the user does not desire to transform qualitative variables, then *Multiple Factor Analysis* is also an option (Bécue-Bertaut & Pagès, 2008).

9 Other coding options are *effect* or *contrast coding* (Alkharusi, 2012).
The second step was to standardize all the variables. Standardization is a common practice in PCA because scale is of high importance to this method. As Lever et al., (2017) explain, “when a small set of variables has a much larger magnitude than others, the components in the PCA analysis are heavily weighted along those variables, while other variables are ignored. Therefore, the PCA simply recovers the values of these high-magnitude variables.” Thus, results could be distorted unless variables with different scales are standardized such that each of them has unit variance\(^{10}\). Given that the current data set had variables that had been dummy coded, and others that were measured in 7-point (Likert) or 10-point (number of interlocutors) scales, and yet others that were composite scores (Richness and Productivity), standardization was necessary.

Finally, the third step was to check for missing data. PCA is more robust when the data have no missing values, as this might skew computation of the covariance matrix of the PCs (Dray & Josse, 2008)\(^{11}\). In the face of this, removing missing values might be preferable, unless it results in severe data loss (i.e., over 30% of datapoints; see Nakagawa & Freckleton, 2008, for a discussion on this topic). In the case at hand, after removing missing values, data loss was 4.2% for the Heritage Language Background dimension, 3.4% for the Contexts of Use dimension, 7.98% for the Interactional Network Dimension, and 3.7% for the Identification, Motivation and Attitude dimension.

After completing three steps above, the \texttt{prcomp()} function (package \texttt{stats}) was used to calculate the PCs via spectral decomposition (i.e., eigendecomposition). This function returned a covariance matrix with the following information:

\(^{10}\) Joliffe & Cadima (2016) warn that if the variables are already on the same scale, standardization is not normally appropriate, as it may actually confound the results.

\(^{11}\) Although there are currently some options to conduct PCA on data sets with missing values (Srebro & Jaakkola, 2003), these require statistically complex and computationally costly methods. In addition, the most widely-used functions available in R (PCA() [FactoMineR] and \texttt{prcomp()/princomp()} [\texttt{stats}]) cannot handle missing values.
1. The standard deviations of the PCs.
2. The matrix of variable loadings; these were the coefficients of the correlation between the original variables and the PCs.
3. The original variables’ means.
4. The original variables’ standard deviations.
5. The coordinates of the observations on the PCs.

Based on this matrix, the \textit{get_eigenvalue} function (package \textit{FactoMineR}) was used to get the eigenvalues for each PC, as well as the percentage of variance explained, and the cumulative percentage of variance explained. Given that an eigenvalue represents the total amount of variance retained by a PC, these three measures basically describe how much of the variance in the data was explained by a particular PC.

It is important to note that the function \textit{prcomp} uses a simple structure \emph{without} rotation. Rotation is not applied to PCA unless there is a good reason to do so because a rotated loading does not yield a principal component \emph{per se}, given that rotated PCs are not orthogonal anymore. The purpose of common rotation methods (e.g., \textit{varimax}, \textit{equamax}, \textit{orthomax}, \textit{quartimax}, \textit{direct oblimin} and \textit{promax}) is to provide a PC structure that is simpler to interpret without sacrificing much variance (Brown, 2009; Jolliffe, 2002). Once the analyses were completed, the results were graphed using the function \textit{fviz_pca_var} (package \textit{factoextra}) to meaningfully visualize the contribution of the different variables to each of the PCs.
Results

The following subsections report PCA results for the four dimensions of the sociolinguistic questionnaire. PCA yielded component groupings that accounted for most of the variance in the data of the current sample of heritage speakers, suggesting that the items in the questionnaire appropriately captured variability in the population that was sampled in this study. For each dimension, the following elements are provided:

1. A table with the original variables analyzed to help better understand the composition of each dimension.

2. A table containing the eigenvalues of the correlation matrix, as well as the percentage of the total variance and cumulative variance explained by each component. This table only shows PCs with eigenvalues higher than 1, that is, those components which account for more variance than the original variables did. This is a widely-used cut-off point to identify the most information-rich components (Peres-Neto & Jackson, 2005; also, see Jollifee, 2002, for an in depth discussion).

3. A two-dimensional correlation circle that illustrates the contribution and correlation of the original variables to PC1 and PC2 via coordinates. Positively correlated variables are grouped together, and negatively correlated variables appear in opposite sides of the plot. According to the color scheme, variables with warmer colors have higher contribution to each PC.

4. A table with numerical values indicating the strength of the correlation and contribution ($r^2$ and $p$-value respectively) of the original variables to each component.
In addition, for each dimension, patterns of variable clusters within a PC1 and PC2 were examined to identify commonalities underlying these clusters and the nature of the PC was logically deduced to provide an interpretation of the descriptive pattern observed (Anderson, Lorinda, Keyvani and Bialystok, 2018; Luk & Bialystok, 2013; Marian et al., 2007).

**Dimension 1: Heritage Language Background**

The variables analyzed in this dimension were the following:

<table>
<thead>
<tr>
<th>Born in the US</th>
<th>Parents’ years in the US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish and English use</td>
<td>Spanish and English AoA</td>
</tr>
<tr>
<td>Spanish and English mean self-ratings</td>
<td>Exposure to Spanish before and after school</td>
</tr>
<tr>
<td>Dominance</td>
<td></td>
</tr>
</tbody>
</table>

Table 2-8 shows that PCs 1–4 had eigenvalues over 1.00. However, just PC1 and PC2 together accounted for more than half of the total variance in the data (51.01%), indicating that these components explained most of the variance in the sample.

<table>
<thead>
<tr>
<th>Component</th>
<th>Eigenvalue</th>
<th>Variance explained %</th>
<th>Cumulative variance %</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>3.54</td>
<td>32.26</td>
<td>32.26</td>
</tr>
<tr>
<td>PC2</td>
<td>2.06</td>
<td>18.75</td>
<td>51.01</td>
</tr>
<tr>
<td>PC3</td>
<td>1.76</td>
<td>16.02</td>
<td>67.04</td>
</tr>
<tr>
<td>PC4</td>
<td>1.09</td>
<td>9.95</td>
<td>76.99</td>
</tr>
</tbody>
</table>

Only components with eigenvalues >1 are shown.

- **PC1**: The variables that contributed the most to this component were *Exposure to Spanish before and after school*, *Age of Acquisition of Spanish and English*, and *Use of Spanish*. Therefore, PC1 could be interpreted to index early experience with...
language, as it describes heritage speakers mostly in terms of their experience with Spanish and English in the early years of their life (see Figure 2-10 and Table 2-9).

- **PC2**: participants’ place of birth (*Born US*), *How many years their parents have been in the US* and, to a lesser extent, *Use of English* were the variables with the highest contribution. Thus, it seems that PC2 could be interpreted as a demographic measure, describing participants mostly in terms of where them and their parents live(d) (see Figure 2-10 and Table 2-9).

Interestingly, the variable *Self-ratings* did not contribute significantly to either component, and although the variable *Dominance* shows a significant correlation with PC1, its significance is much lower than that of other variables (Table 2-9). Thus, the results of the PCA on the *Heritage Language Background* dimension suggest that the heritage speakers in this sample might be best differentiated in terms of their early experience with the heritage and non-heritage language, as these factors explain the most variance among individuals. Secondarily, demographics also play a role in differentiating participants, while other variables such as, dominance and, most of all, self-rated proficiency have less explanatory power.
Table 2-9. Correlation of variables with PCs – Heritage Language Background

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure before school</td>
<td>0.88</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Use Spanish</td>
<td>-0.80</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Exposure after school</td>
<td>0.70</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>AoA Spanish</td>
<td>0.69</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>AoA English</td>
<td>-0.63</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Dominance</td>
<td>0.45</td>
<td>0.02</td>
</tr>
<tr>
<td>Born in the U.S.</td>
<td>0.78</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Parents’ years in the U.S.</td>
<td>0.72</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Use of English</td>
<td>0.64</td>
<td>&gt; .001</td>
</tr>
</tbody>
</table>
The two first PCs had eigenvalues over 1.00 and, together, accounted for 71.35% of the total variance in the data (Table 2-10).

**Table 2-10.** Eigenvalues of the correlation matrix – Contexts of Use

<table>
<thead>
<tr>
<th>Principal component</th>
<th>Eigenvalue</th>
<th>Variance %</th>
<th>Cumulative variance %</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>3.18</td>
<td>53.01</td>
<td>53.01</td>
</tr>
<tr>
<td>PC2</td>
<td>1.10</td>
<td>18.3</td>
<td>71.35</td>
</tr>
</tbody>
</table>

Only components with eigenvalues >1 are shown.

- **PC1 and PC2:** The variables with the greatest contribution were Richness of Free time, Hobbies and School, while Richness of Home and Work contributed the most to PC2 (see Figure 2-11 and Table 2-11).

The raw questionnaire data show that Free time, Hobbies and School are the contexts in which language use is most variable, showing a mean Richness that ranges between 0.56 and 0.45 (out of 1). Use at Home and Work is less variable, while Home is the richest context (0.70/1), Work is the poorest (0.40/1). A possible interpretation of these results is that use of the heritage language in contexts with low variability is more constant and similar across speakers, as would be the case with Spanish being spoken the most at home and the least at Work. However, use of the heritage language in contexts with high variability (Free time, Hobbies and School) fluctuates more and diverges across speakers.
Thus, given that most heritage speakers use the heritage language to a similar degree at home and at work, it is use in contexts with high variability that explains the most variance among individuals. The results of the PCA on the *Contexts of Use* dimension indicate that we can best characterize the heritage speakers in this sample in terms of the Richness of Contexts of Use with high variability.

**Figure 2-11.** Contribution of variables to PCs—Contexts of Use

![Figure 2-11.](image)

**Table 2-11. Correlation of variables with PCs—Contexts of Use**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richness Free time</td>
<td>0.93</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Richness Hobbies</td>
<td>0.88</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Richness School</td>
<td>0.73</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Overall activity</td>
<td>0.73</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Richness Home</td>
<td>0.90</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Richness Work</td>
<td>-0.41</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Interactional Network

The analysis of this dimension was split into: (a) The Productivity of the Network and (b) The Bilingualism of the Network.

Productivity of the Network

A total of four variables were analyzed in this sub-dimension:

<table>
<thead>
<tr>
<th>Productivity of immediate family</th>
<th>Productivity of extended family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productivity of friends</td>
<td>Productivity of acquaintances</td>
</tr>
</tbody>
</table>

Table 2-12 shows that the two first PCs had eigenvalues over 1.00 and, together, accounted for 82.56% of the total variance in the data.

<table>
<thead>
<tr>
<th>Principal component</th>
<th>Eigenvalue</th>
<th>Variance %</th>
<th>Cumulative variance %</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>2.34</td>
<td>58.66</td>
<td>58.66</td>
</tr>
<tr>
<td>PC2</td>
<td>1.00</td>
<td>23.90</td>
<td>82.56</td>
</tr>
</tbody>
</table>

Only components with eigenvalues >1 are shown.

- **PC1 and PC2**: The variables with the greatest contribution to this component were Productivity Immediate Family and Extended Family Networks, while Productivity Acquaintances and Friends Networks contributed the most to PC2 (see Figure 2-12 and Table 2-13).
The variables analyzed in this sub-dimension were the following:

**Bilingualism of the Network**

- Bilingualism of immediate family
- Bilingualism of extended family
- Bilingualism of friends
- Bilingualism of acquaintances

**Table 2-13. Correlation of variables with PCs – Productivity of Network**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1 Productivity immediate family</td>
<td>0.84</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>PC1 Productivity extended family</td>
<td>0.79</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>PC2 Productivity acquaintances</td>
<td>0.66</td>
<td>&gt; .01</td>
</tr>
<tr>
<td>PC2 Productivity friends</td>
<td>-0.51</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Figure 2-12. Contribution of variables to PCs – Productivity of Network**

The variables analyzed in this sub-dimension were the following:

<table>
<thead>
<tr>
<th>Bilingualism of immediate family</th>
<th>Bilingualism of extended family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilingualism of friends</td>
<td>Bilingualism of acquaintances</td>
</tr>
</tbody>
</table>
The two first PCs had eigenvalues over 1.00, and, together, accounted for 78.73% of the total variance in the data (Table 2-14)

**Table 2-14. Eigenvalues of the correlation matrix – Bilingualism of Network**

<table>
<thead>
<tr>
<th>Principal component</th>
<th>Eigenvalue</th>
<th>Variance %</th>
<th>Cumulative variance %</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>1.68</td>
<td>42.23</td>
<td>42.2</td>
</tr>
<tr>
<td>PC2</td>
<td>1.46</td>
<td>36.50</td>
<td>78.73</td>
</tr>
</tbody>
</table>

Only components with eigenvalues >1 are shown

- **PC1 and PC2**: The variables with the greatest contribution were *Bilingualism Immediate Family* and *Extended Family Networks*, while *Bilingualism of Acquaintances* and *Friends Networks* contributed the most to PC2 (see Figure 2-13 and Table 2-15).

**Figure 2-13. Contribution of variables to PCs – Bilingualism of Network**
According to the raw questionnaire data, the *Immediate Family* and *Extended Family Networks* are the largest ones in size, with a mean of 7.03 and 5.08 interlocutors respectively (in a scale of 1-10). On the contrary, the *Friends* and *Acquaintances Networks* are smaller, with 5.00 and 3.92 interlocutors on average. Thus, we could interpret these results as indicating that the largest networks have the greatest impact on heritage speakers’ linguistic experience. The results of the PCA on the *Interactional Network* dimension suggest that the heritage speakers in this sample can be best differentiated in terms of the Productivity and Bilingualism of their family interactional networks, as these larger networks explain the most variance among individuals.

### Attitude, Motivation and Identity

Seven variables were analyzed in this dimension:

<table>
<thead>
<tr>
<th>Identification with Spanish</th>
<th>Identification with English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude towards Spanish</td>
<td>(Dis)advantage to Spanish</td>
</tr>
<tr>
<td>Motivation to learn Spanish</td>
<td>Spanish course years</td>
</tr>
<tr>
<td>Opinion of own Spanish</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1 Bilingualism immediate family</td>
<td>0.87</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>PC1 Bilingualism extended family</td>
<td>0.76</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>PC1 Bilingualism friends</td>
<td>-0.59</td>
<td>0.01</td>
</tr>
<tr>
<td>PC2 Bilingualism acquaintances</td>
<td>0.90</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>PC2 Bilingualism friends</td>
<td>0.63</td>
<td>&gt; .01</td>
</tr>
</tbody>
</table>
Table 2-16 shows that the two first PCs had eigenvalues over 1.00, and, together, accounted for 64.20% of the total variance in the data.

<table>
<thead>
<tr>
<th>Principal component</th>
<th>Eigenvalue</th>
<th>Variance %</th>
<th>Cumulative variance %</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>1.75</td>
<td>25.02</td>
<td>25.02</td>
</tr>
<tr>
<td>PC2</td>
<td>1.58</td>
<td>22.61</td>
<td>47.63</td>
</tr>
<tr>
<td>PC3</td>
<td>1.15</td>
<td>16.56</td>
<td>64.20</td>
</tr>
</tbody>
</table>

Only components with eigenvalues >1 are shown

- **PC1**: The variables that contributed the most to this component were Identification with Spanish, Identification with English, and Spanish course years. Thus, PC1 could be interpreted to index identification because describes heritage speakers mostly in terms of their identification with the heritage and non-heritage language, and considers aspects of their education (see Figure 2-14 and Table 2-17).

- **PC2**: Participants’ Opinion of own Spanish, Attitude, and (Dis)advantage to Spanish were the variables with the highest contribution. Thus, it seems that PC2 serves as a measure of participants’ perception of the heritage language (see Figure 2-14 and Table 2-17).

Interestingly, the variable Motivation did not contribute significantly to either component. In all, the results of the PCA on the Attitude, Motivation and Identity dimension indicate that we can best characterize the heritage speakers in this sample in terms of their identification with the heritage and non-heritage language, as these factors explain
the most variance among individuals. In addition, the educational experience in the heritage language is also a factor with high explanatory power.

**Figure 2-14.** Contribution of variables to PCs – Attitude, Motivation & Identity

![Diagram showing contribution of variables to PCs](image)

**Table 2-17.** Correlation of variables with PCs - Attitude, Motivation & Identity

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification with Spanish</td>
<td>0.74</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Identification with English</td>
<td>0.73</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Spanish course years</td>
<td>0.71</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Opinion of own Spanish</td>
<td>-0.67</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>Attitude</td>
<td>0.65</td>
<td>&gt; .001</td>
</tr>
<tr>
<td>(Dis)advantage to Spanish</td>
<td>0.59</td>
<td>&gt; .001</td>
</tr>
</tbody>
</table>
Correlation with behavioral measures

A final step was taken to test the validity and explanatory power of the PCs extracted. The first PC of each dimension was correlated with heritage speakers’ performance in the behavioral task Bilingual Picture Naming (BPN), which participants had completed both in Spanish and English (see the Participants subsection). While other proficiency measures had also been collected (i.e., Verbal fluency and Elicited Imitation Task), the PCA results were correlated only with the BPN data for two reasons: (1) these showed the most variability among participants (RT range= 898.77, accuracy range= 28.00); (2) given the complexity of PCA data, it is best practice to reduce the number of correlations with behavioral measures in order simplify the results reported (Anderson et al., 2018; Luk & Bialystok, 2013; Marian et al., 2007).

Table 2-18 shows the correlation of all four dimensions with the BPN task; results indicate that all dimensions correlate significantly with the task in at least one language, and that the correlations are stronger (higher Pearson’s $r^{12}$) and more numerous for the Spanish version.

<table>
<thead>
<tr>
<th>Heritage Language Background</th>
<th>Contexts of Use</th>
<th>Interactional Network</th>
<th>Attitude, Motivation &amp; Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPN Sp</td>
<td>$r$</td>
<td>95% CI</td>
<td>$r$</td>
</tr>
<tr>
<td>-.33</td>
<td>[-.68, .42]</td>
<td>-.36</td>
<td>[-.65, .01]</td>
</tr>
<tr>
<td>BPN En</td>
<td>-.15</td>
<td>[.27, .52]</td>
<td>-.14</td>
</tr>
</tbody>
</table>

$^{12}$ An $r^2$ higher than .30 is commonly considered a strong correlation (Akoglu, 2018).
**Discussion**

Here, it was argued that discrepancies in the results reported by studies with heritage speakers might, at least in part, be caused by the way heritage samples are characterized. Thus, the goal of this study was to address a common problem faced by researchers using LHQs to characterize and differentiate bilingual populations according to their linguistic experience: LHQs yield a high number of highly correlated variables that are hard to disentangle. This issue becomes even more challenging in the case of highly variable bilinguals such as heritage speakers. Here, bilingual experience was operationalized into four core dimensions and PCA was used as an empirical method to explore these dimensions in order to maximize the explanatory power of LHQ data. In the following subsection, I discuss in more depth the results of the PCA analysis and how they go in line previous research on heritage populations. For the sake of clarity, Table 2-19 below provides a summary of the results obtained:

<table>
<thead>
<tr>
<th>Dimension</th>
<th>PCA Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heritage Language Background</strong></td>
<td>participants can be best differentiated in terms of their early experience with the heritage and non-heritage language and, secondarily, by their demographics.</td>
</tr>
<tr>
<td><strong>Contexts of Use</strong></td>
<td>use of the heritage language in contexts with high variability explains the most variance among individuals.</td>
</tr>
<tr>
<td><strong>Interactional Network</strong></td>
<td>participants can be best characterized in terms of the productivity and bilingualism of their family interactional networks.</td>
</tr>
<tr>
<td><strong>Attitude, Motivation &amp; Identity</strong></td>
<td>identification with the heritage and non-heritage language are the factors that explain the most variance among individuals, followed by educational experience in the heritage language.</td>
</tr>
</tbody>
</table>
Correlates of heritage linguistic experience

*Heritage Language Background*

Within this dimension, PCA indicated that early experience with the heritage and non-heritage language was the factor that explained most of the variance in the data set and best differentiated the individuals in the heritage sample. This was followed closely by the demographics of participants and their parents. The impact of early linguistic experience on bilingualism has been discussed in studies that address the role of age on language acquisition (e.g., Abrahamsson & Hyltenstam, 2009; Birdsong & Molis, 2001; DeKeyser, 2000; Jia & Aaronson, 2003; Johnson & Newport, 1989; Montrul, 2008; Yeni-Komshian, Flege, & Liu, 2000; also see Montrul, 2015 for an in-depth review). Most heritage speakers are simultaneous bilinguals or early sequential bilinguals (i.e., child L2 learners). In the sample at hand, while there was very little variability in the AoA of the heritage language ($M= 0.26$, $SD= 0.77$, range= 2), the AoA of the non-heritage language presented much higher variation ($M= 3.61$, $SD= 2.38$, range= 8). Regarding early exposure, the data indicated that participants had been exposed to Spanish (i.e., the heritage language) significantly earlier than English (i.e., the non-heritage language) ($t(40) = -7.10$, $p < .001$). Only 3 individuals reported having exposure to both languages before starting school, the majority were only exposed to the heritage language. Exposure to the non-heritage language at home increased after starting school (N= 9/33, where 9 individuals reported exposure to both languages). Despite acquiring Spanish earlier, slightly over half of the participants reported considering themselves English dominant (N= 19/33).

The PCA results suggest that early experience with the heritage and non-heritage language is impacted by demographics. As explained in the Method section (see subsection *Demographic and linguistic characteristics*), most participants in the current
sample were born in New Mexico (N= 26/33) and the remaining 6 were born in Mexico and moved to the US as infants (mean age of immigration 2.50 years old, SD= 1.38, range= 4). Participants’ parents were also mostly born in New Mexico (N= 50/66) and the remaining 16 were born in Mexico and then moved to Albuquerque. Importantly, there was significant variability in the number of years the Mexico-born parents had been living in the US (M= 25.44 years, SD= 5.98, range= 22.5). To illustrate the relevance of this variability in the development of heritage grammars, an individual whose parents were born in Mexico and have not lived in the US for long will most likely primarily use the heritage language at home, which might in turn delay exposure to the non-heritage language. By comparison, a heritage speaker with one of the parents born in the US will be exposed and use the non-heritage earlier and more frequently.

In addition, it was also observed that Self-Ratings did not contribute significantly to PC1 or PC2 in the Heritage Language Background dimension. The fact that self-ratings can be unreliable has been discussed in several studies. For instance, Flege, Frieda, and Nozawa (1997), and Flege, Frieda, Walley and Randazza (1998) showed that bilinguals’ self-ratings were less predictive of L2 performance than experience-related variables (see also: MacIntyre et al., 1997). More recently, a study by Tomoschuk et al., (2019) tested the validity of self-ratings both within and between different bilingual groups by comparing self-rated proficiency to scores on the Multilingual Naming Test (MINT; Gollan, Weissberger, Runnqvist, Montoya & Cera, 2012). These measures were examined in two groups of bilinguals matched in language-pairing and proficiency level: the first was a group of Chinese-English bilinguals, and the second a group of Spanish-English bilinguals. Results indicated that, when compared to the MINT results, self-ratings showed significant discrepancies both between and within groups even though participants had been matched across several variables. For example, English-dominant
Chinese-English bilinguals scored like other groups in the MINT; however, they had overall lower self-ratings than English-dominant Spanish-English bilinguals. In addition, they also had lower self-ratings that their Chinese-dominant Chinese-English peers.

The authors suggest that a possible explanation for this pattern of results is that different participant groups have different frames of reference that they use to evaluate their language proficiency (2019: 28); this entails that using self-ratings of proficiency as a measure to compare participant groups with different language is highly unreliable and that objective proficiency measures (e.g., picture naming, proficiency interviews, MINT, etc.) are better as they maximize classification accuracy and consistency across studies. As mentioned before (see Introduction), using self-ratings with heritage populations is problematic because these individuals are oftentimes the target of a double stigma: not only are they speakers of a minority language but, in many cases, they have learned varieties of the heritage language that are stigmatized in the home country (Valdés, 1997; Poplack & Torres Cacoullos, 2015). Thus, it is common to observe that they rate their linguistic skills overall lower than other bilingual populations, which makes self-ratings largely unreliable in this population (Giancaspro, 2017; Perez-Cortes, 2016; Perez-Cortes, Putnam & Sanchez, 2019).

In all, the examination of the data from the sociolinguistic questionnaire together with the results of the PCA analysis suggests the findings of the present study go in line with previous research, underscoring the relevance of early linguistic experience and its interaction with demographics, as well as highlighting the fact the linguistic background of bilinguals crucially affects their self-rating judgments, which can make this measure unreliable for the study of heritage populations.
**Contexts of Use**

Within this dimension, PCA indicated that the relative richness of contexts of use with high variability (Richness of Free time, Hobbies and School) was the factor that explained most of the variance in the data, and best differentiated the individuals in the current heritage sample. The factor with the second highest explanatory power was the relative richness of contexts of use with low variability (Richness of Home and Work).

The impact of the context of use in bilingualism is a topic that has received increasing attention over the last decades. Of relevance is the Adaptive Control Hypothesis (ACH; Green & Abutalebi, 2013), according to which bilinguals’ language control abilities are shaped by the contexts in which they use their languages (see also Wu & Thierry, 2013). Green and Abutalebi identify three types of contexts of use based on the degree of control needed over language selection and the proficiency required by the interlocutors (i.e., single-language, dual-language, and dense code-switching). In the data at hand, contexts of use with high variability can be equated with dual-language contexts where both languages are used with different speakers within the same context or maybe even with dense code-switching contexts where speakers interleave their languages in the course of a single utterance (2013: 518). In the same way, contexts of use with low variability can be equated with single-language contexts, where there is a one-language, one-context separation.

The present study supports the ACH by providing evidence that variability in the contexts of use of the heritage language can be used as a fine-grained measure of differentiation within heritage populations, as shown in the results of the PCA analysis. In addition, it can also be a significant predictor of language performance. Figure 2-15 below shows that heritage speakers who used the heritage language more frequently in
contexts with high variability had lower RTs when naming pictures in Spanish ($r = .36$, CI[-.65, .18]).

![Figure 2-15. Correlation between PC1 – Contexts of Use and Spanish Picture Naming](image)

To conclude, the results of this dimension highlight the relevance of the contexts of use, and extends these findings to heritage populations.

**Interactional Network**

This dimension was split into the *Productivity* and *Bilingualism* of the different network groups. For both, PCA indicated that the productivity and bilingualism of the family networks (immediate and extended) were the factors that explained most of the variance in the data, and best differentiated the individuals in the current heritage sample. The productivity and bilingualism of the friends and acquaintances networks were the factors with the second highest explanatory power.
An ample body of literature suggests that the interactional network of heritage speakers is vital for the acquisition and maintenance of the heritage language. The most commonly studied variables are the size (i.e., number of interlocutors) and density (i.e., a network is dense if people interact regularly) of the network. For instance, in a study with Chinese–English and Spanish–English heritage speakers, Gollan, Starr and Ferreira (2015) compared how picture naming scores and frequency of use of the heritage language correlated with participants’ interactional network size. Results indicated that the size of the network accounted for unique variance in picture naming independently, and above, frequency of use. Similarly, Gibbons and Ramirez (2004) investigated the effects of network density (close network/family vs. distant network/acquaintances) on maintenance of the heritage language. They found that correlations between network density and proficiency measures (fluency, accent, idiomaticity and mixing) were statistically robust and, importantly, stronger for the closer network (for similar results see Gathercole & Thomas, 2009; Kenji & D’Andrea, 1992; Oh, 2003; Place & Hoff, 2011; Rivera-Mills, 2012; Tsai et al., 2012).

In the present study, I contribute the measures of productivity and bilingualism of the network to the current literature. In line with other studies, the results of the PCA analysis indicated that it was the productivity and bilingualism of the interactional networks with the highest number of interlocutors (i.e., immediate and extended family) that best differentiated participants within the current sample of heritage speakers. Given that early experience is primarily provided by family members, having a large, productive and highly bilingual family network is crucial to promote bilingualism. This

13 Gibbons and Ramirez’ (2004) statistical models showed that the estimate for network size accounted for more variance than the estimate for frequency of use. In addition, the fact that there was not an interaction between these variables lead the authors to conclude that the effect of network size was independent from that of frequency of use.
fact is supported by the data in Figure 2-16, which shows that heritage speakers with highly productive and bilingual family networks had lower RTs when naming pictures both in English and Spanish (English: $r = .35$, CI[-.70, .11]; Spanish: $r = .55$, CI[-.80, -.12]).

![Figure 2-16. Correlation between PC1 – Interactional Network and Picture Naming](image)

To summarize, these results emphasize the relevance of the family interactional network and extend these findings to new variables to be considered in future studies.

**Attitude, Motivation, and Identity**

Within this dimension, PCA indicated that identification with the heritage language was the factor that explained most of the variance in the data and best differentiated the individuals in the current heritage sample. The factor with the second highest explanatory power was perception of the heritage language.

Psychosocial variables such as identification and language attitudes have long been shown to be play an important role in the development of the heritage language
(Bhatt; 2009; Bell-Corales, 2006; Carreira, 2004; Chinen & Tucker, 2005; Kondo-Brown, 2005; also, see Kagan, 2012, and Pascual y Cabo, 2016, for in-depth reviews). For instance, language attitudes contribute to the construction of heritage speakers’ identity and how they perceive and interact with their community of practice which, in turn, modulates heritage language outcomes. In a study of Korean heritage speakers, Cho (2000) found that participants’ proficiency was modulated language attitudes, such that individuals with more positive attitudes had higher proficiency levels. Along similar lines, Torres, Estremera and Mohamed (2019) reported that “psychosocial and biographical variables predicted participants’ performance differentially across linguistic knowledge tasks” (p.17) in heritage speakers of Spanish. They found that while language attitudes significantly correlated with scores in an Elicited Imitation Task, factors such as motivation and willingness to communicate better correlated with performance in a Grammaticality Judgement Task.

The PCA results obtained in the present study support these results in two ways: first, they show that identification-related affective factors are empirically significant descriptors of the heritage sample analyzed. Second, these factors were significantly correlated with language performance. There was a significant negative correlation between Spanish BPN and PC1 of dimension *Attitude, Motivation & Identity*. Higher PC1 scores indicate that a participant felt more identified with the heritage language. According to Figure 2-17, individuals with higher PC1 scores had lower RTs when naming pictures in Spanish.
In addition to identification with the heritage language, the educational experience in the heritage language was a factor with high explanatory power in PC1. It is a well-studied fact that literacy in the heritage language has a significant positive impact on heritage speakers’ proficiency (Kondo-Brown 2009; Kupisch, 2013; Montrul, 2013; Montrul et al., 2008; Pires & Rothman 2009; Rothman 2007; Zaretsky & Bar-Shalom, 2010). For instance, Swender, Martín, Rivera-Martinez and Kagan (2014) measured the proficiency levels of heritage speakers of Russian and Spanish using the American Council on the Teaching of Foreign Languages (ACTFL) oral proficiency test. They
investigated whether participants’ years of formal education in Russian or Spanish were correlated with the scores they received in the test. Their results showed a significant correlation between formal education in the heritage language and the scores ratings for each proficiency level. Importantly, the correlation was stronger at higher proficiency levels. Similarly, in a series of studies, Kupisch, Lein, Barton, Schröder, Stangen and Stöhr (2013) and Kupisch, Akpınar & Stöhr (2013) examined heritage speakers of French living in Germany who had attended a French school. They investigated several linguistic dimensions (i.e., adjective placement, article use, gender assignment and agreement, choice of prepositions, and voice onset time) and found no significant differences between these heritage speakers and a control group of bilinguals neither in production nor in comprehension. Thus, the fact that in the present study years of learning (PC1) was significantly correlated with better performance in the Spanish BPN task, further supports the findings of studies such as the ones discussed above.

Finally, it was observed that the factor Motivation did not contribute significantly to PC1 or PC2 in the Attitude, Motivation, and Identification dimension. This was an unexpected finding given that motivation has been shown to positively impact heritage language outcomes (see Ducar, 2012, and Lu and Li, 2008 for reviews). This might be due to the fact that there was only one question in the sociolinguistic questionnaire that asked about motivation. There is evidence that single-question self-assessments can be highly unreliable, especially in bilinguals (Delgado, Goggin & Ellis, 1999; Paap & Greenberg, 2013). Having several questions capturing the same construct in multiple ways is more effective than a single question (Anderson et al., 2018:261; see also Dunning, Heath & Suls, 2004).

In all, the examination of the data from the sociolinguistic questionnaire in light of the results of the PCA analysis brings attention to the relevance of affective factors,
showing their potential for the differentiation of individuals within heritage populations. In addition, the results underscore the importance of assessing questionnaire constructs via multiple questions in order to increase the reliability of the results obtained with bilingual groups.

Achieving a more meaningful characterization of heritage speakers and their grammars.

Scholars within the field of heritage language studies have made substantial efforts to produce a definition of the term *heritage speaker* that is flexible yet concrete enough to have descriptive power. The difficulty in defining the term is that heritage populations and their grammars are extremely diverse, presenting high inter- and intra-category variability than other bilingual groups. A thorough survey of the literature shows that definitions of the term heritage speaker pivot around two main dimensions: proficiency and cultural identification. Definitions based on proficiency, also often include notions of language dominance (Treffers-Daller 2011). For example, Valdés (2000: 38) defines a heritage speaker as “someone […] who speaks or merely understands the heritage language and who is to some degree bilingual in English and the heritage language.” Similarly, Benmamoun, Montrul and Polinsky (2013: 260) say that heritage speakers are “asymmetrical bilinguals who learned the heritage language as an L1 in childhood, but who, as adults, are dominant in a different language” (see also: Beaudrie and Fairclough 2012; Carreira & Kagan 2011; Polinsky and Kagan 2007; Rothman 2009; Valdés 2001; Wiley and Valdés 2000). In contrast, definitions based on cultural identification describe heritage speakers in terms of their cultural ties to the heritage language. For instance, McCarty et. al. (1997: 20) define heritage speakers as “individuals who are members of a community with linguistic roots in a language other than English”. Along similar lines,
Carreira (2004: 1) proposes that heritage speakers are individuals that “have identity and/or linguistic needs [...] that relate to their family background” (see also: Cho, 2000; Fishman, 2001; Van Deusen-Scholl, 2003).

The variety and lack of concreteness of these definitions raises the possibility that defining heritage speakers based solely on their linguistic knowledge and cultural identification may not be very useful. Indeed, Zyzik (2016) proposes a definition that makes use of prototype theory. Prototypes have fuzzy boundaries such that some members of a category are more central to it (i.e., prototypical) than others. Zyzik suggests that a prototype of the heritage speaker be built around the following scalar attributes:

- Early exposure to the heritage language in the home.
- Proficiency in the heritage language
- Degree of bilingualism
- Dominance in a language other than the heritage language.
- Ethnic/Cultural connection to the heritage language.
- Implicit knowledge of the heritage language.

These attributes may collectively serve to profile the prototypical heritage speakers and, at the same time, recognize that membership in this category is a matter of degree (p. 28).

In all, the incredible variability of heritage grammars and the extraordinary diversity of heritage speakers’ linguistic experience challenge the idea that heritage speakers can be meaningfully categorized solely in terms of their grammars at all. The terms heritage language, heritage grammar and heritage speaker actually refer to a wide range of realities, defined not so much through characteristics of the constructs
themselves as through those of their environments and circumstances. For this reason, a better understanding of the socially-situated process of heritage language acquisition and use would allow for more meaningful characterizations of heritage grammars and their speakers. Here, I propose a usage-based dynamic approach according to which acquisition and use of the heritage language vary along the following dimensions:

- **When**: this dimension is comprised by factors related to the age of acquisition and years of use of the heritage language.

- **Where**: this dimension is comprised by factors related to the contexts within which the heritage language is used and acquired. It also includes aspects pertaining to literacy.

- **Who**: this dimension is comprised by factors related to the interactional networks within which the heritage language is used and acquired.

Under this approach, a heritage speaker is but an individual who acquires a heritage language and the linguistic knowledge resulting from their acquisition process is a product of variation along the dimensions explained above. Variation along the when, where and who dimensions of the heritage language will determine the variety of constructions and vocabulary that heritage speakers are exposed, directly impacting their grammatical representations and processing strategies.

### Conclusion

There is a need for scholars in the field of heritage language research to provide more detailed and meaningful characterizations of their heritage speaker samples (Benmamoun, Montrul & Polinsky, 2013; Meisel, 2019). More nuanced descriptions and
differentiations of the different heritage populations studied will directly impact the interpretation and extrapolation of results across studies, as well as the theories derived from them. Here, I have presented PCA as an empirical method to generate evidence-based characterizations of heritage speakers by examining which language-experience related factors best explain the variability inherent to a sample. To complement this, I have also proposed a theoretical approach from which to dynamically understand how variation along language-experience related factors shape heritage grammars. Hopefully, these tools can contribute towards a more unified and accurate account of language acquisition and processing in heritage contexts.
CHAPTER 3: INVESTIGATING SENSITIVITY TO THE LEXICAL AND STRUCTURAL CONDITIONING OF MOOD SELECTION

Introduction

Heritage grammars have unique properties that set them apart from those of other speaker groups. Traditionally, differences between heritage and non-heritage grammars have been interpreted as the product of simplification, attrition or incomplete acquisition affecting comprehension and production (Imada, Zhang, Cheour, Taulu, Ahonen & Kuhl, 2006; Kuhl & Rivera-Gaxiola, 2008; Montrul, 2007, 2009, 2015; Polinsky, 2011; Silva-Corvalán, 1994). Studies assessing the grammatical knowledge of speakers of Spanish as a heritage language, have largely focused on the Spanish subjunctive mood, likely because prescriptively mastering the subjunctive is considered a significant landmark of acquisition, and an indicator of high proficiency (Collentine, 2003). These studies have concluded, almost unanimously, that heritage speakers’ knowledge of the Spanish subjunctive is non-native-like (Montrul, 2007, 2009, et seq.; Montrul & Perpiñán, 2011; Silva-Corvalán, 1994). However, a series of recent studies have challenged these deficit-oriented perspectives (Ortega, 2013) providing evidence that, while different from that of non-heritage speakers, heritage speakers’ linguistic knowledge is by no means deficient (Bookhammer, 2013; Giancaspro, 2017; LaCasse, 2018; Perez-Cortes, 2016, 2020).

The present study addresses this discrepancy by integrating corpus data into the design of psycholinguistic experiments created to examine knowledge of the Spanish subjunctive in heritage speakers who live in a long-standing bilingual community in Albuquerque, New Mexico. Their performance was compared to a sample of bilingual native speakers of Mexican Spanish; these participants were bilingual but considered Spanish to be their dominant language, so this group is referred to as Spanish-dominant
bilinguals henceforth. To compare both groups, I first, investigated heritage speakers’ sensitivity to (1) the lexical and (2) structural factors that condition mood selection in Spanish, both in comprehension and production (Torres Cacoullos et al., 2017; Poplack et al., 2018). Participants completed an auditory pupillometry experiment designed to investigate their processing of the subjunctive during online comprehension (Experiment 1) and an elicited production experiment that examined their production of the subjunctive (Experiment 2). I also explored other lexical factors (i.e., morphological regularity and lexical frequency) that might modulate speakers’ sensitivity to the factors that condition mood selection (Giancaspro, 2017; Perez-Cortes, 2020). In addition, participants had completed a sociolinguistic questionnaire and a battery of behavioral tasks that measured their proficiency in both languages. In the case of heritage speakers, given that the conditions under which the heritage language is acquired impact heritage grammar outcomes (Torres, Estremera & Mohamed 2019; see Aalberse & Muysken 2013 for a review), taking into consideration not just proficiency, but also sociolinguistic factors will allow scholars to achieve a better understanding of their knowledge of the Spanish subjunctive. Based on the PCA analysis conducted in Chapter 2 the predictive power of three sociolinguistic factors beyond proficiency was examined in the heritage sample: use of the heritage language, language entropy and identification with the heritage language.

In all, by investigating comprehension and production, together with linguistic and sociolinguistic variables, this study seeks to provide a more holistic understanding of heritage grammars and to contribute to a theory of language acquisition and processing in heritage language contexts that has greater explanatory adequacy. In the chapter at hand, I will first describe and motivate the methodology used. Next, I will present the research questions that this study aims to answer. Then, I will describe the
methods and results separately for each group. Finally, I provide a discussion of the findings.

**What is pupillometry?**

Pupillometry is an online data collection technique employed in language science research. Psychological and neurological work over the past several decades has shown that the pupillary response is linked not only to changes in ambient luminance, but also to aspects of the sympathetic nervous system (Goldwater, 1972), as well as the *locus coeruleus* and the norepinephric system (LC-NE; Aston-Jones & Cohen, 2005). The LC-NE has been associated with aspects of memory retrieval (Attar, Schneps & Pomplum, 2013), selective attention (Foote & Morrison, 1987), and arousal (Bradshaw, 1967); for example, when selective attention is engaged, either due to cognitive demands or attentional requirements, the pupil dilates in response to these stimuli due to its connection to the LC-NE.

Recently, pupillometry has been applied to the study of a variety of language-related processes, such as effortful speech processing (Kuchinsky et al., 2013), lexical retrieval (Schmidtke, 2014), bilingual cognate facilitation (Guasch, Ferré & Haro, 2017), and the processing of language mixing (Byers-Heinlein, Morin-Lessard & Lew-Williams, 2017), highlighting its sensitivity to language processing phenomena (see Schmidtke, 2018 for an in-depth review). For instance, at the word level, it has been shown that lower frequency words produce higher pupillary dilation both in monolinguals and bilinguals (Conklin & Pellicer-Sanchez, 2016; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). At the sentence level, Engelhardt, Ferreira, and Patsenko (2010) showed that
garden-path sentences were harder to understand, thus producing higher pupillary dilation, when the prosody of the sentence mismatched its syntactic structure.

One of the main advantages of pupillometry over other psycholinguistic methods, is that it is well-suited for the presentation of complex auditory stimuli. As such, this methodology affords greater flexibility in experimental design and allows for more ecologically valid stimuli. Pupillometry is ideal for auditory stimuli because it requires no visual information to be present, allowing for less restricted—and therefore more naturalistic—materials that are presented as a continuous speech stream. These advantages notwithstanding, a few methodological controls—screen luminance (brightness) and ambient lighting—must be employed. While the former is not an issue with auditory stimuli (the screen can be kept constant), the latter is highly important, as pupil dilation due to changes in ambient lighting can obscure stimulus-evoked pupillary responses (Schmidtke, 2018).

Most experimental work within the field of heritage language research to this day has made use of off-line tasks such as acceptability judgements, sentence conjunction, oral and written elicitation, morphology recognition and preference tasks (Giancaspro, 2017; Mikulski, 2010; Montrul, 2007, 2009, 2011, et seq.; Montrul & Perpiñán, 2011; Perez-Cortes, 2016, 2020). Nevertheless, as discussed in Chapter 1 subsection Task effects, there is evidence that these tasks may not be entirely appropriate for heritage speakers, given that this population has been shown to underperform on written tasks and on explicit tasks that require the use of metalinguistic knowledge (Alarcón, 2011; Bowles, 2011; Méndez, Rothman & Slabakova, 2015; Montrul, 2010; Montrul, Foote & Perpiñán, 2008; Montrul, 2010; Montrul, Davidson, De la Fuente & Foote, 2014). As explained before, these findings are not surprising if one considers heritage speakers’ acquisitional background; these speakers learn the heritage language mostly aurally and exhibit high
variability in their literacy levels in the heritage language. For this reason, the first experiment in this dissertation employs auditory *pupillometry*, an on-line, implicit, aural eye-tracking method that is more natural, and therefore more ecologically valid, in relation to heritage speakers’ linguistic experience.

**Research questions**

As explained in Chapter 1, prescriptive notions of mood have a crucial shortcoming: because they are not based on the observation of natural language use, they cannot explain variation patterns that break away from prescriptive approaches. For instance, the binary distinction between obligatory and variable contexts of use cannot account for the fact that there is variation in subjunctive selection among monolingual varieties of Spanish (Gallego & Alonso-Marks, 2014; Morales, 1989; Serrano, 2004; Pousada, Poplack & Sankoff, 1982). Likewise, normative characterizations based on the purported semantic or pragmatic functions of the subjunctive can readily be contested with examples from corpora of natural language (Torres Cacoullos et al., 2017). These facts strongly suggest that prescriptive notions might not be suitable to study the Spanish subjunctive.

The two experiments described below draw from two recent variationist studies: Poplack et al., (2018) and Torres Cacoullos et al., (2017) (see Chapter 1, subsection *Variationist characterizations of the subjunctive* for an in-depth review). The first study provides a pan-Romance characterization of the subjunctive mood based on naturally-produced speech, while the second examines the Spanish subjunctive diachronically across three written texts. These two studies have proved crucial to scholars’ current understanding of the subjunctive because they have debunked the notion that mood selection is based solely on semantic and syntactic factors (Bolinger, 1974; Foster, 1973;
Harris, 1974; Murphy, 2008; Terrell & Hooper, 1974; Villalta, 2000; Wheelock, 2000). Instead, Poplack and Torres Cacoullos demonstrated that the Spanish subjunctive mood is primarily *lexically* conditioned across Romance languages. That is, the nature of the governing verb (i.e., a subjunctive selecting verb) in the main clause predominantly determines the presence of a verb with subjunctive morphology in the subsequent subordinate clause (e.g., $\text{Quiero}_{\text{gov}}$ que $\text{venga}_{\text{subj}}$ a la fiesta). In addition, their results suggest that Spanish is the only Romance language in which structural factors still condition subjunctive selection, albeit to a lesser degree than lexical factors.

Thus, in Spanish the probability of a subordinate verb in the subjunctive mood is a function of the *subjunctive selection rate* inherent to the governor and the structural elements present in the utterance. As such, governors can be classified according to their variability: Non-variable governors categorically select the subjunctive 100% of the time, while variable governors select the subjunctive anywhere between 3% to 90% of the time, for instance. Crucially, however, when a variable governor is negated, the subordinate verb following it is significantly more likely to appear in the subjunctive (e.g., No $\text{creo.}_{\text{gov}}$ que $\text{venga}_{\text{subj}}$ a la fiesta), to the point that many of these governors also become Non-variable. Therefore, negated governors have a high tendency to also select the subjunctive.

The present study focuses on Non-variable governors and negated governors, using an auditory pupillometry experiment and an elicited production experiment to examine what factors condition subjunctive selection in comprehension and production:

(1) *Lexical conditioning of mood selection:* Comparison 1 employs Non-variable governors to investigate whether participants are sensitive to the
conditioning that Non-variable governors categorically require the presence of a subjunctive verb in embedded clauses.

(2) **Structural conditioning of mood selection:** Comparison 2 employs negated governors to investigate whether participants are sensitive to the conditioning that negated governors tend to require the presence of a subjunctive verb in embedded clauses.

The two experiments described here were designed to address the following research questions:

**Research question 1:** *Are heritage speakers who live in a long-standing bilingual community sensitive to both the lexical and structural factors that condition mood selection in Spanish?*

Recent studies (Bookhammer, 2013; Giancaspro, 2017; LaCasse, 2018; Perez-Cortes, 2016) have shown that heritage speakers can display robust and systematic knowledge of the subjunctive mood comprehension and production. Thus, it is predicted that the current sample of heritage speakers will be sensitive to the lexical and structural conditioning of mood selection in Spanish, showing a processing cost in contexts where the indicative is used but the subjunctive is expected.

**Research question 2:** *What other linguistic aspects modulate the degree of sensitivity displayed by heritage speakers?*

Previous research has shown that factors such as the lexical frequency of the governor and the morphological regularity of the subordinate verb (e.g., *voy/venga* v.s. *vaya/venga*)...
quiero/quieras; ‘to go’, ‘to want’) might influence the degree to which heritage and L2 speakers and are sensitive to mood selection (Giancaspro, 2017; Gudmestad, 2006, 2010, 2012; Lubbers Quesada, 1998; Perez-Cortes, 2020). Thus, it is predicted that regularity and lexical frequency will also modulate sensitivity to mood selection in the current sample of heritage speakers.

Research question 3: Do sociolinguistic factors also modulate heritage speakers’ sensitivity to mood selection?

There is evidence that sociolinguistic factors can modulate mood selection in monolinguals and L2 speakers, albeit to a lesser extent than linguistic factors (Gudmestad, 2010, 2012, 2013). Thus, it is predicted that use of the heritage language, language entropy and identification with the heritage language will also modulate sensitivity to mood selection in the current sample of heritage speakers.

Research question 4: Are there asymmetries in heritage speakers’ sensitivity to the conditioning of mood selection in comprehension vs. production?

Studies have shown that heritage speakers often display an asymmetry between their comprehension and production abilities (Mikulski, 2006; Mikulski & Elola, 2013; Montrul, 2009, 2011; Perez-Cortes, 2016; Putnam & Sanchez, 2013; Rothman, Pascual y Cabo & Lingwall, 2012; Sherkina-Lieber, Pérez-Leroux & Johns, 2011; Valdes Kroff, Villegas and Dussias, under review). However, it is possible that these asymmetries might be, at least in part, caused by the methodology and the group of heritage speakers investigated (LaCasse, 2018). Given that the heritage community in Albuquerque is well-established and bilingualism has a strong tendency to be maintained across generations.
it is predicted that the potential asymmetry between comprehension and production of the subjunctive will be absent or minimized in the current sample of heritage speakers.

**Research question 5:** Does heritage speakers’ sensitivity to the lexical and structural factors that constrain mood diverge from that of Spanish-dominant bilinguals of the target language? If so, what is the nature of this divergence?

In the present study, the performance of heritage will be compared to a group of non-Spanish-dominant bilinguals from Mexico as each speaker group listens to sentences while their pupil dilation is being recorded. Three alternatives are hypothesized:

1. Divergence between both groups will be categorical, such that only the non-heritage group will display sensitivity.

2. Divergence between both groups will be graded. Because heritage speakers experience with Spanish is not as extensive and varied as that of the non-heritage group, the degree of sensitivity might be lower for heritage speakers.

If hypothesis 2 is true, there is the possibility that sensitivity to other lexical aspects

**Group 1: Heritage Speakers**

**Experiment 1: Auditory Pupillometry Study**

**Participants.** The same 33 heritage speakers that completed the sociolinguistic questionnaire in Chapter 2 participated in the pupillometry experiment in this study.

**Materials.** The same materials were used for the auditory pupillometry and elicited production experiments. In the following subsections, I first discuss the auditory pupillometry experiment, describing the experimental comparisons and explaining how
materials were distributed across conditions and counterbalanced. After this, I provide a
detailed description of how materials were designed and their characteristics. Finally, I
proceed to describe the elicited production experiment in the same manner. All materials
can be found in Appendix A.

**Experimental comparisons and counterbalancing.** The experiment consisted of
two experimental comparisons, each of which contrasted a condition with a subordinate
verb in the *subjunctive* relative to a condition with a subordinate verb in the *indicative*. In
both cases, it was predicted that the condition with a subordinate verb in the indicative
would be more costly to process (i.e., cause larger pupillary dilation) on the grounds that
Non-variable governors and negated governors favor the subjunctive. All experimental
sentences were paired with a preceding context and a comprehension question. The
target region for analyses was the subordinate verb right after the complementizer *que*
(‘that’).

**Experimental comparison 1: Lexical Conditioning**

1. Condition *Non-variable Subjunctive* consisted of a Non-variable governor (in
   bold in Table 3-1) followed by a subordinate verb in the subjunctive (in italics
   in Table 3-1).

2. Condition *Non-variable Indicative* was comprised of a Non-variable governor (in
   bold in Table 3-1) followed by a subordinate verb in the indicative (in italics in
   Table 3-1).
### Table 3-1. Auditory Pupillometry Comparison 1—*Lexical Conditioning*

<table>
<thead>
<tr>
<th>Original token</th>
<th>Contextual sentence</th>
<th>Condition</th>
<th>Target Sentence</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entonces ¿quieres que <em>apague</em> la luz?</td>
<td>Mi hijo tiene 7 años, pero todavía le tiene miedo a la oscuridad y por eso duerme con varias luces prendidas. Sin embargo, creo que eso va a cambiar hoy.</td>
<td><strong>Non-variable Subjunctive</strong></td>
<td>De hecho, mi hijo <em>quiere</em> que <em>apague</em> todas las luces.</td>
<td>¿Él quiere las luces apagadas?</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Non-variable Indicative</strong></td>
<td>De hecho, mi hijo <em>quiere</em> que <em>apago</em> todas las luces.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘So, do you want me to <em>turn</em> off the light?’</td>
</tr>
<tr>
<td>‘My son is 7 years old, but he is still afraid of the dark, so he sleeps with the lights on. However, I think that might change tonight.’</td>
</tr>
<tr>
<td>‘In fact, my son <em>wants</em> me to <em>turn</em> off all the lights.’</td>
</tr>
<tr>
<td>Does he want the lights off?</td>
</tr>
</tbody>
</table>

### Experimental comparison 2: *Structural Conditioning*

1. Condition *Negated Subjunctive* consisted of a negated governor (in bold in Table 3-2) followed by a subordinate verb in the subjunctive (in italics in Table 3-2).

2. Condition *Negated Indicative* was comprised of a negated governor (in bold in Table 3-2) followed by a subordinate verb in the indicative (in italics in Table 3-2).
For the *Lexical Conditioning* and the *Structural Conditioning* comparisons, 128 experimental sentences were created, 64 in each condition. This resulted in a total of 256 experimental sentences. In addition, 64 filler sentences and 8 practice sentences were also created, resulting in a total of 328 items. To counterbalance these materials, two 200-item lists were created, each with:

- 8 practice items-identical across lists.
- 128 experimental sentences-64 Non-variable governors (32 by condition) and 64 negated governors (32 by condition).
- 64 fillers-identical across lists.

Items were counterbalanced across both lists such that a single participant did not see the same sentence in more than one condition. In addition, all items within a list were
presented in a randomized fashion to avoid order effects. Finally, the lists were split in half (i.e., 100 items each) and administered in two different sessions to make the experiment shorter and avoid participants getting fatigued, as this can decrease the quality of the data recorded. The flowchart below (Figure 3-1) shows the distribution of items per list:

**Figure 3-1.** Stimulus distribution per list – Auditory Pupillometry

Design of materials. The stimuli were derived from 1153 tokens extracted by Torres Cacoullos et al., (2017) from 36 of the 108 interviews that constitute the *Corpus Sociolingüístico de la Ciudad de México*, (CSCM; Martín Butragueño & Lastra, 2011, 2012, 2015); this subset of the CSCM amounted to about 400,000 words.

For the *Lexical Conditioning* comparison, 14 Non-variable governors were chosen to examine sensitivity to the lexical conditioning of subjunctive selection. All governors belonged to the volitional or desiderative semantic classes, their frequency ranged from 1-72 per 400,000 words in the CSCM subset (M= 14.07, SD= 23.86), and their rate of
subjunctive selection was 100% (see Table 3-3). When designing the materials, the main goal was to present participants with ecologically valid stimuli that mirrored natural utterances produced by speakers in their everyday lives. The original tokens were edited as little as possible in order to keep them true to real life usage while still having them conform to experimental rigorousness. Two strategies were used to control the materials: (1) The number of sentences per governor was made proportional to its frequency in the CSCM, and (2) The tokens were edited to control for a series of morphosyntactic features, as well as length.

1. **Proportion of sentences**

The total number of experimental sentences per governor was calculated proportionally to the governor’s frequency in the CSCM, such that governors that were more frequent in the corpus had correspondingly more sentences in the experiment. The proportion of experimental sentences per governor was calculated by multiplying the frequency of each governor by 60 (the minimum necessary number of stimuli per condition for a pupillometry experiment), dividing it by the sum of the frequency of all governors (197), and rounding up the result. For instance, ‘to say’, *decir*: \((72 \times 60)/197 = 21.93\), rounded to 22.00. This resulted in a total of 64 experimental sentences, which were then created in Condition *Non-variable Subjunctive* and Condition *Non-variable Indicative*, producing a final total of 128 items.
2. Morphosyntactic features

- **Tense and regularity**: Spanish presents stem changes in some verbs that differ between the present and the preterit. The tense of all tokens was controlled by transforming the verbs into the present tense. Still, even in the present tense, some verbs were irregular, that is, they had stem changes according to person and number (e.g., Indicative 3rd = *trae* vs Subjunctive 3rd sg. = *traiga*, ‘to bring’). However, this feature was not edited.

- **Subject expression**: Spanish is a pro-drop language in which subjects are often omitted. The subjects of all tokens were controlled by modifying them such that the subject of the governor was always explicit, and the subject of the subordinate verb was always implicit.

### Table 3-3. Non-variable governors used in Comparison 1—*Lexical Conditioning*

<table>
<thead>
<tr>
<th>Governor</th>
<th>Semantic</th>
<th>Sub. Rate</th>
<th>Freq.</th>
<th>Prop. of sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>autorizar</td>
<td>volition</td>
<td>100%</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>intentar</td>
<td>desiderative</td>
<td>100%</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>requerir</td>
<td>volition</td>
<td>100%</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>conseguir</td>
<td>volition</td>
<td>100%</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>insistir</td>
<td>volition</td>
<td>100%</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>permitir</td>
<td>volition</td>
<td>100%</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>preferir</td>
<td>volition</td>
<td>100%</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>mandar</td>
<td>volition</td>
<td>100%</td>
<td>3</td>
<td>1.00</td>
</tr>
<tr>
<td>evitar</td>
<td>volition</td>
<td>100%</td>
<td>5</td>
<td>1.00</td>
</tr>
<tr>
<td>pedir</td>
<td>volition</td>
<td>100%</td>
<td>9</td>
<td>3.00</td>
</tr>
<tr>
<td>dejar</td>
<td>volition</td>
<td>100%</td>
<td>12</td>
<td>4.00</td>
</tr>
<tr>
<td>hacer</td>
<td>volition</td>
<td>100%</td>
<td>19</td>
<td>6.00</td>
</tr>
<tr>
<td>querer</td>
<td>desiderative</td>
<td>100%</td>
<td>66</td>
<td>20.00</td>
</tr>
<tr>
<td>decir</td>
<td>volition</td>
<td>100%</td>
<td>72</td>
<td>22.00</td>
</tr>
</tbody>
</table>

**Total: 64.00**
• **Length:** a sentence-initial preamble and sentence-final adjunct was added to each stimulus to make them more natural. All preambles consisted of one or two words and all adjuncts consisted of three. All sentences had a length of 10 words.

• **Person and Number:** the grammatical number of all tokens was controlled by converting them into the 3rd person singular or plural, as these were the most common throughout all tokens. The number of the subordinate verb remained as in the original token unless it needed to be modified for semantic congruency. Figures 3-2 and 3-3 below show the distribution of both Non-variable governors and subordinate verbs. There was a total of 45 subordinate verbs distributed across all 14 governors; 42 of these subordinate verbs were regular. Subordinate verbs were coded as regular or irregular following the guidelines in Gudmestad (2006) and Collentine (1995); irregular verbs were limited to those whose present indicative and present subjunctive *stems* were different (e.g., *tener*, ‘to have’: *tienes*).

![Figure 3-2. Person and Number distribution of Non-variable governors](image)

**Figure 3-2. Person and Number distribution of Non-variable governors**

**Comparison 1 - Lexical Conditioning**
For the Structural Conditioning comparison, 10 negated governors were chosen to examine sensitivity to the structural conditioning subjunctive selection. All governors belonged to the evaluative semantic class, their frequency ranged from 2-33 per 400,000 words in the CSCM subset (M= 7.8.00, SD= 10.76), and their subjunctive selection rate ranged from 3%-85%. However, under negation, they behaved categorically, selecting the subjunctive 100% of the time. As explained above, the materials were edited and controlled in the using two strategies:

1. **Proportion of sentences**

The proportion of sentences per governor was calculated in the same way as that of Non-variable governors, and also resulted in a total of 64 experimental sentences, which produced a final total of 128 items when created in Condition 1–Negated Subjunctive and Condition 2–Negated Indicative (see Table 3-4).
Table 3-4. Negated variable governors used in Comparison 2 – Structural Conditioning

<table>
<thead>
<tr>
<th>Governor</th>
<th>Semantic</th>
<th>Subj. Rate</th>
<th>Subj. Rate + Neg</th>
<th>Freq.</th>
<th>Prop. of sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>no lamento</td>
<td>evaluative</td>
<td>80%</td>
<td>100%</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>no dudo</td>
<td>evaluative</td>
<td>64%</td>
<td>100%</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>no admito</td>
<td>evaluative</td>
<td>72%</td>
<td>100%</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>no recomiendo</td>
<td>evaluative</td>
<td>50%</td>
<td>100%</td>
<td>2</td>
<td>2.00</td>
</tr>
<tr>
<td>no acepto</td>
<td>evaluative</td>
<td>46%</td>
<td>100%</td>
<td>3</td>
<td>3.00</td>
</tr>
<tr>
<td>no necesito</td>
<td>evaluative</td>
<td>85%</td>
<td>100%</td>
<td>3</td>
<td>3.00</td>
</tr>
<tr>
<td>no pienso</td>
<td>evaluative</td>
<td>33%</td>
<td>100%</td>
<td>3</td>
<td>3.00</td>
</tr>
<tr>
<td>no espero</td>
<td>evaluative</td>
<td>11%</td>
<td>100%</td>
<td>6</td>
<td>5.00</td>
</tr>
<tr>
<td>no creo</td>
<td>evaluative</td>
<td>77%</td>
<td>100%</td>
<td>22</td>
<td>19.00</td>
</tr>
<tr>
<td>no supongo</td>
<td>evaluative</td>
<td>3%</td>
<td>100%</td>
<td>33</td>
<td>26.00</td>
</tr>
</tbody>
</table>

Total = 64

2. Morphosyntactic features

Tokens were also controlled for tense, form regularity, subject expression, length, person and number. Concerning person and number, Figures 3-4 and 3-5 below show the distribution of negated governors and subordinate verbs. There was a total of 16 subordinate verbs distributed across all 10 governors, 12 of these subordinate verbs were regular.
Figure 3-4. Person and Number distribution of negated governors
Comparison 2–Structural Conditioning

Figure 3-5. Person and Number distribution of subordinate verbs
Comparison 2–Structural Conditioning
Finally, a set of 64 filler items was created which consisted of 32 items containing governors with low rates of subjunctive selection (i.e., less than 50%), as in example (3.1), and 32 containing subjunctive verbs within conditional clauses, as in example (3.2):

(3.1) Por eso, mi hermana decide que tiene que dejar algunos.
‘For that reason, my sister decides that she has to leave some behind.’

(3.2) Si su hija lo entendiera, ella querría algo menos caro.
‘If his daughter understood this, she would want something cheaper.’

All filler governors belonged to semantic classes other than volitional, desiderative or evaluative, their frequency ranged from 2-109 per 400,000 words in the CSCM subset (M= 16.50, SD= 33.89), and their subjunctive selection rate ranged from 3%-50% (M= 26.58, SD= 31.95). Fillers were also edited and controlled for proportion of sentences (see Table 3-5) and morphosyntactic features. Regarding the person and number of governors and subordinate verbs in filler sentences, most governors were in the 3rd person singular; there were a total of 16 verbs distributed across all 10 fillers, 12 of these subordinate verbs were regular.

<table>
<thead>
<tr>
<th>Governor</th>
<th>Subj. Rate</th>
<th>Frequency</th>
<th>Proportion of sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>decidir</td>
<td>50%</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>avisar</td>
<td>25%</td>
<td>3</td>
<td>1.00</td>
</tr>
<tr>
<td>sentir</td>
<td>9%</td>
<td>3</td>
<td>1.00</td>
</tr>
<tr>
<td>asegurar</td>
<td>40%</td>
<td>5</td>
<td>2.00</td>
</tr>
<tr>
<td>entender</td>
<td>14%</td>
<td>24</td>
<td>6.00</td>
</tr>
<tr>
<td>imaginar</td>
<td>4%</td>
<td>26</td>
<td>7.00</td>
</tr>
<tr>
<td>ver</td>
<td>12%</td>
<td>28</td>
<td>8.00</td>
</tr>
<tr>
<td>saber</td>
<td>6%</td>
<td>41</td>
<td>11.00</td>
</tr>
<tr>
<td>creer</td>
<td>3%</td>
<td>109</td>
<td>27.00</td>
</tr>
</tbody>
</table>

Total = 64
Experimental design. A female native speaker from Ciudad de México checked all stimuli for coherence and naturalness. Stimuli where recorded by this same female assisted by the experimenter in a sound-attenuated booth using a two-channel, high-resolution USBPre 2 audio interface and a Shure SM35 head-worn cardioid condenser headset microphone. After being recorded, all stimuli were annotated in Praat (version 6.1.11) in the following manner (see Figure 3-6 for a visual illustration):

- Boundaries were set to delimit each context and target sentence.
- Four tiers were created to annotate each stimulus for: item number, condition, governor, type (sentence or context).
- A fifth point-tier was created to mark the onset of the subordinate verb (target region) in each target sentence.

Figure 3-6. Praat annotation scheme

NVI: Condition Non-variable Indicative; NVS: Condition Non-variable Subjunctive

Praat scripts were used to extract and splice all target sentences, as well as to normalize their intensity. After the target sentences were recorded, they were extracted individually and the carrier sentence (in bold in Table 3-6) was split from the subordinate
sentence (in *italics* in Table 3-6) for each condition (see Figure 3-6). Then, each subordinate sentence was spliced into the carrier extracted from the condition in which the governor was followed by a subordinate verb in the subjunctive (i.e., Condition *Non-variable Subjunctive* and *Negated Subjunctive*).

| Table 3-6. Splicing of experimental sentences – Auditory Pupillometry |
|---------------------------------------------------------------|-----------------------------|
| Carrier | Subordinate |
| Al final, su jefe autoriza | que trabaje_{SUBJ} los fines de semana |
| | que trabaja_{IND} los fines de semana. |

*Translation: ‘Her boss authorizes that she works during the weekends.’*

Splicing the target sentences was crucial because the conditions in which the governor was followed by a subordinate verb in the indicative are unattested in the natural speech of monolingual speakers (i.e., they are *ungrammatical*). Therefore, there was a possibility that the native speaker who recorded the target sentences might have unknowingly produced them with a different prosodic modulation caused by the upcoming ungrammaticalicity in the subordinate verb. In comparison, the conditions in which the governor was followed by a subordinate verb in the subjunctive are attested in natural speech (i.e., *grammatical*) and would have been read neutrally. Thus, by means of splicing I made sure that the carrier phrase was identical for each condition, ensuring that the prosodic information leading up the subordinate verb (i.e., target verb) did not create different expectations for each condition.

Finally, the use of pupillometry entails carries with it certain requirements in the design and presentation of stimuli that must be met. First, the overall luminance or brightness of the display screen must be kept constant. To achieve this, participants were instructed to fixate on a drawing of an ear presented in the middle of the screen. This
image was present throughout the duration of the stimulus period, ensuring there were no changes in luminance throughout the trial. Second, given that gaze position has been found to systematically influence the calculation of pupil size (Gagli, Hawelka & Hutzler, 2011), participants were instructed to not move their eyes away from the image presented in the middle of the screen. Third, pupil size also changes over longer periods of time due to natural fluctuations in overall attention levels and sustained processing (Sirois & Brisson, 2014:680; Schmidtke, 2018: 531). To address this, the pupillary response was time-locked to the onset of the target subordinate verb and extended for 3000 ms, this time period served as the primary interest period for analysis (i.e., target period).

In addition to these requirements, a 1000 ms neutral period where no audio was playing, was presented before each stimulus to ensure that there was sufficient time for the pupillary response to stabilize before the next sentence began (i.e., baseline period). Lastly, because the pupillary response is most robust when the participant must actively engage with the stimuli and perform an additional task, participants were asked to use the keyboard to answer yes/no comprehension questions about the stimulus they had just heard. In all, each experimental trial consisted of the following:

1. The baseline period, which lasted 1000 ms, had no audio and displayed a drawing of an ear in a central 300- by-300-pixel interest area.
2. The stimulus, which began playing after the 1000 ms baseline period; the drawing of the ear remained on-screen such that the luminance did not change.
3. The target period, beginning at the onset of the target subordinate verb and extending 3000 ms; the drawing of the ear remained on-screen.
4. The comprehension question task, beginning after the 3000 ms offset period.
Figure 3-7 below shows the trial design. Note that for fillers, which did not include a target subordinate verb, the image of the ear remained on-screen for 2000 ms after the end of the sentence finished playing, at which point the comprehension question appeared.

**Figure 3-7. Trial design – Auditory Pupillometry**

---

**Procedure.** The laboratory where the experimental sessions took place consisted of a windowless room, which ensured that the only light source—an overhead light fixture—remained constant at all times. All data were collected using an Eyelink Portable Duo eye-tracker (SR Research), which recorded participants’ pupil size and fixation information at 1000 Hz in head-stabilized mode using the right eye pupil and corneal reflection. After informed consent was obtained, participants were instructed to sit in a chair and rest their heads in a chin rest to ensure stability during recording, as significant movements of the head have been found to affect pupil size (Gagl, Hawelka & Hutzler, 2011). After the task was explained and questions were answered, a nine-point calibration was performed such that the average error was no greater than 0.5 degrees and the maximum error was no greater than 1.0 degrees. Each session then began with a practice consisting of four items, during which the participant could ask questions or re-do the
practice if necessary. Once the actual task began, the experiment was broken down into two blocks of 50 items each with a break in between which participants used to rest their eyes; they were encouraged to take as many breaks as they needed throughout the session.

Once the eye-tracking task finished, participants completed the verbal fluency and elicited imitation tasks in session one, and the picture naming task and sociolinguistic interview in session two. Participants were paid separately at the end of each session at a rate of 20 USD per hour, and a short debriefing was held at the end of the second session where they informed of the general goals of the study and were asked about their overall thoughts on the various tasks.

**Data pre-processing and cleaning.** Initial data extraction took place in SR Research DataViewer (v. 3.2.48) using a ‘Time Course (Binning) Analysis.’ The interest period was set to start 500 ms pre-stimulus onset and end 3000 ms post-stimulus onset (i.e., at the end of the target period). Data were binned into 20 ms time bins, and the following variables were selected to be extracted for each bin:

- The average right-eye pupil size across all non-blink samples.
- The average right-eye x- and y-gaze position across all non-blink samples.
- The proportion of the total number of samples that were in a blink event.
- The proportion of the total number of samples that were in a saccade event.
- The proportion of the total number of samples that fell in the central 300-by-300 pixel interest area.
- The proportion of the total number of samples that fell outside of the central 300-by-300 pixel interest area.
• The proportion of the total number of samples that fell off-screen.

Data from four participants was removed because they did not complete all two sessions of the study, and data from five more participants was removed because they had more than 50% missing data. In total, data was extracted for 33 participants. After extraction, R Core Team (2020) was used to pre-process the data. The average pupil size in the baseline period was used to baseline-correct the pupil size during the 3000 ms target period. Baseline-correction was performed by subtracting each trial’s average pupil size in the baseline period from the average pupil size in each bin of the target period (see Van Rij, Hendriks, Van Rijn, Baayen, & Wood, 2019:3-6). Trials in the baseline or target periods where more than 25% of the samples occurred in a blink or saccade event or outside of the central interest were excluded from further analysis, resulting in 15% of trials being excluded, which is not abnormally high for pupillary data (Schmidtke, 2014). Finally, all trials that included an inaccurate response to the comprehension question were also excluded, resulting in the removal of 9% of the remaining data.

**Data analysis.** Separate models were created to examine:

(1) The *Lexical Conditioning* comparison (*Non-variable Subjunctive vs. Non-variable Indicative*).

(2) The *Structural Conditioning* comparison (*Negated Subjunctive vs. Negated Indicative*).

(3) The effects of linguistic variables on the *Lexical Conditioning* and the *Structural Conditioning* comparisons.

   a. Morphological regularity.
b. Lexical frequency.

(4) The effects of sociolinguistic variables on the *Lexical Conditioning* and the *Structural Conditioning* comparison.

   a. Proficiency
   b. Use of the heritage language.
   c. Language entropy.
   d. Identification with the heritage language.

Pupil size was modeled as a time-dependent variable using generalized additive mixed-effects models (GAMMs) through the `mgcv` package (v. 1.8-31) in R. Analyses were carried out following the recommendations found in Van Rij, Hendriks, Van Rijn, Baayen, and Wood (2019) and Wieling (2018), such that all models were specified to:

1. Use a scaled $t$-distribution to account for the non-normal distribution of the data.
2. Include an autoregressive error model for the residuals (AR(1)) to correct for autocorrelation among datapoints, helping to adjust the confidence of the models’ estimates.
3. Include a smooth term capturing the interaction between the $x$- and $y$-gaze positions, used to account for the effects of gaze position on pupil size over time (Gagl et al., 2011)
4. Include by-participant and by-item nonlinear factor smooths, used to capture potentially non-linear trends over time (penalization m=1, basis dimension k=10).

Thus, the models described below only differed in the variables and interactions included, as will be discussed in turn.
Lexical Conditioning and Structural Conditioning comparisons. The two main models for the Lexical Conditioning comparison (conditions Non-variable Subjunctive vs. Non-variable Indicative) and the Structural Conditioning comparison (conditions Negated Subjunctive vs. Negated Indicative) predicted the baseline-corrected pupil size by a binary difference smooth. When using binary difference smooths, one smooth represents the reference level of a factor, and the other represents the difference between the reference and the other level of the factor. Thus, binary smooths make it possible to infer whether the difference between the levels of a factor is significant directly from the model output (Wieling, 2018: 93). In the two main models, the difference between conditions was examined by using the factor levels Non-variable Subjunctive and Negated Subjunctive as reference levels for comparison in their respective models.

Linguistic Variables. First, the models fitted to examine the effect of the morphological regularity of the subordinate verb on the Lexical Conditioning and the Structural Conditioning comparisons also used binary difference smooths. To model interactions between categorical factors (i.e., regularity and condition) in GAMMs, separate models had to be specified to investigate the effect of regularity (a 2-level factor) on both levels of condition independently. Thus, a total of four models with different reference levels were created:\textsuperscript{14}:

- For the NVS model, Non-variable Subjunctive Irregular was the reference level.
- For the NVI model, Non-variable Indicative Regular was the reference level.
- For the Neg-S model, Negated Subjunctive Irregular was the reference level.

\textsuperscript{14} NVS/NVI: Non-variable Subjunctive/Indicative; Neg-S/Neg-I: Negated Subjunctive/Indicative;
• For the Neg-I model, *Negated Indicative Regular* was the reference level.

Second, the models fitted to examine the effect of the lexical frequency of the governor on the *Lexical Conditioning* and the *Structural Conditioning* comparisons used tensor product smooths, rather than binary difference smooths. To explore the effects of a continuous variable (i.e., *frequency*) on a categorical variable and (i.e., *condition*) in GAMMs, the former needed to be modeled as an interaction with time, given that time-series data are being analyzed. The resulting non-linear interaction accounted for changes in pupil size caused by *frequency* on *condition* over time.

**Sociolinguistic Variables.** The models created to examine the effect of *proficiency*, use of the heritage language, language entropy and identification with the heritage language on the *Lexical Conditioning* and the *Structural Conditioning* comparisons also used tensor product smooths to model the effect of these continuous variables on *condition* over time.

**Results**

**Lexical and Structural Conditioning comparisons.** In GAMMS, visualization is crucial for meaningfully interpreting the non-linear terms included in a model (Van Rij et al., 2019: 11; Wieling, 2018). In the present study, all plots were produced using the *itsadug* package (v. 2.3).

• Lexical Conditioning Figure 3-8 below shows the fitted smooths for the NVS and NVI conditions extracted from the main model of the *Lexical Conditioning* comparison. Descriptively, condition *Non-variable Indicative* elicited a larger pupillary response (higher magnitude and duration)
starting at bin 17 (340 ms into the target period) in comparison to condition Non-variable Subjunctive:

**Figure 3-8.** Pupillary Response Smooths based on Fitted Values
Lexical Conditioning – Non-variable Governors

The output of the model (Table 3-7) shows that the non-linear pattern for the difference between the two conditions (in bold) was highly significant:

**Table 3-7.** Main model output – *Lexical Conditioning*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVS)</td>
<td>-30.41</td>
<td>4.25</td>
<td>-7.15</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVS)</td>
<td>8.89</td>
<td>11.09</td>
<td>4.07</td>
<td>&lt; .001</td>
</tr>
<tr>
<td><strong>Diff. NVI minus NVS</strong></td>
<td><strong>10.44</strong></td>
<td><strong>13.14</strong></td>
<td><strong>2.81</strong></td>
<td><strong>&lt; .001</strong></td>
</tr>
<tr>
<td>Gaze x by Gaze y</td>
<td>27.58</td>
<td>28.81</td>
<td>300.27</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>183.17</td>
<td>330.00</td>
<td>1.54</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>242.57</td>
<td>610.00</td>
<td>1.15</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .09; Deviance Explained: 9.56%
• Structural Conditioning Figure 3-9 below shows the fitted smooths for the Neg-S and Neg-I conditions extracted from the main model of the *Structural Conditioning* comparison. Descriptively, condition *Negated Indicative* elicited a larger pupillary response starting at bin 40 (800 ms into the target period) in comparison to condition *Negated Subjunctive*.

**Figure 3-9. Pupillary Response Smooths based on Fitted Values**

*Structural Conditioning – Negated Governors*

The output of the model (Table 3-8) shows that the non-linear pattern for the difference between the two conditions (*in bold*) was significant.

**Table 3-8. Main model output – Structural Conditioning**

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVS)</td>
<td>-29.20</td>
<td>4.17</td>
<td>-7.01</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVS)</td>
<td>7.77</td>
<td>9.83</td>
<td>5.00</td>
<td>&lt; .001</td>
</tr>
<tr>
<td><strong>Diff. Neg-I minus Neg-S</strong></td>
<td>2.89</td>
<td>3.44</td>
<td>2.71</td>
<td>.02</td>
</tr>
<tr>
<td>Gaze x by Gaze y</td>
<td>27.96</td>
<td>28.89</td>
<td>378.77</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>202.36</td>
<td>329.00</td>
<td>2.06</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>193.70</td>
<td>579.00</td>
<td>0.98</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>
In summary, the results presented so far suggest that heritage speakers who live in a long-standing bilingual community are sensitive both to the lexical and structural factors that condition mood selection in Spanish during online comprehension. In both cases, the differences between conditions were robust, peaking early and being maintained throughout the target period. Interestingly, it was observed that the magnitude of this effect was stronger for those conditions manipulating lexical factors (i.e. Non-variable governors).

In addition to Lexical and Structural Conditioning, two other linguistic variables that might modulate heritage speakers’ sensitivity were examined. Based on previous research, it was hypothesized that the lexical frequency of the governor and the morphological regularity of the subordinate verb might influence heritage speakers’ sensitivity to mood selection (Giancaspro, 2017; Gudmestad, 2006, 2010, 2012; Lubbers Quesada, 1998; Perez-Cortes, 2020). The results for the analyses of these factors are described in turn below.

**Linguistic Variables**

**Morphological Regularity.** As explained, four separate models were created to investigate the effect of the morphological regularity of the subordinate verb on the Lexical Conditioning (Non-variable governors) and Structural Conditioning (Negated governors) comparisons (tables with the models’ outputs can be found in Appendix A, Tables A-D). Figure 3-10 below shows the fitted smooths for the Non-variable Subjunctive Regular/Irregular and Non-variable Indicative Regular/Irregular conditions. Descriptively, condition Non-variable Indicative Regular elicited a larger pupillary response starting at bin
60 (1200 ms into the target period) in comparison to condition *Non-variable Indicative Irregular*. Likewise, condition *Non-variable Subjunctive Regular* elicited a larger pupillary response starting at bin 32 (640 ms into the target period) in comparison to condition *Non-variable Subjunctive Irregular*. As can be observed, the magnitude and duration of the difference in pupil-size is higher for the *Non-variable Indicative* conditions.

**Figure 3-10.** Pupillary Response Smooths based on Fitted Values
Regularity – Non-variable Governors

The output of the NVS and NVI models (Appendix A, Tables A and B) showed that the non-linear pattern for the difference between the regular and irregular subordinate verbs was not significant. However, both models showed a significant interaction between *condition* and *regularity* (NVS: $F = 4.20$, $p < .001$; NVI: $F = 9.31$, $p < .001$); visualization of the fitted smooths (Figure 3-10) and the results of the NVI model suggest that this interaction was largely driven by the significant difference between
conditions Non-variable Indicative Regular and Non-variable Subjunctive Regular \( (F = 4.16, p < .001; \text{in bold in Table B in Appendix A}) \), where the effect of regularity was stronger for the Non-variable Indicative condition.

Fitted smooths for the Negated Subjunctive Regular/Irregular and Negated Indicative Regular/Irregular conditions are visualized in Figure 3-11. Descriptively, condition Negated Indicative Regular elicited a larger pupillary response starting at bin 30 (600 ms into the target period) in comparison to condition Non-variable Indicative Irregular. However, condition Negated Subjunctive Regular did not consistently elicit a larger pupillary response throughout the target period in comparison to condition Negated Subjunctive Irregular. That is, the magnitude and duration of the difference in pupil-size is higher for the Negated Indicative conditions.

**Figure 3-11.** Pupillary Response Smooths based on Fitted Values
Regularity – Negated Governors

For models, Neg-S and Neg-I (Appendix A, Tables C and D) the non-linear pattern for the difference between the regular and irregular subordinate verbs was not significant for either condition, and neither was there an interaction between condition and regularity. The only significant effect captured in model Neg-S was for the difference between
conditions Negated Indicative Irregular and Negated Subjunctive Irregular \( (F = 3.13, p = .02) \) where the pupillary response was higher for the indicative condition.

To sum up, the results of the analyses conducted indicate that morphological regularity does modulate heritage speakers’ sensitivity to mood selection. Overall, regular subordinate verbs are harder to process than irregular subordinate verbs; this difficulty is exacerbated when the governor requires the subjunctive, but the subordinate verb is in the indicative. Importantly, the findings suggest that the effect of morphological regularity is selective in the heritage sample, as it is only observed for Non-variable governors.

**Lexical Frequency.** As mentioned, models with tensor product smooths were created to investigate the effect of the lexical frequency of the governor on the Lexical Conditioning (Non-variable governors) and Structural Conditioning (Negated governors) comparisons. In this case, visualization was essential to understand the results because it was used to determine significance: to visualize a two-dimensional pattern such as the effect of frequency on condition over time, contour plots based on the model’s predictions were plotted. Here, the color bands represent the range of values of the dependent variable, in this case the difference in pupillary response between conditions (i.e., NVI minus NVS, or Neg-I minus Neg-S), the closer to red, the greater the difference. The highlighted areas in the plot indicate where there was a significant effect of the independent variable (y-axis) on the dependent variable, and the x-axis shows the time course of the effect.

- For the *Lexical Conditioning* comparison (see model output in Appendix A, Table E), Figure 3-12 below shows a strong significant difference in
pupillary response between conditions for high frequency governors starting approximately at bin 70 (140 ms into the target period). The difference smooths showed that both the magnitude and duration of the frequency effect were greater for higher frequency governors, although the effect for low frequency governors peaked earlier.

**Figure 3-12.** Difference in Pupillary Response based on Fitted Values

- For the *Structural Conditioning* comparison (see model output in Appendix A, Table F), Figure 3-13 shows there was a strong significant difference in pupillary response between conditions for both high frequency governors and low frequency governors (highlights the area around the middle of the y-axis), starting approximately at bin 50 (100 ms into the target period). The difference smooths showed that, once again, the magnitude of the frequency effect was greater for higher frequency governors, although the effect for low frequency governors peaked earlier and lasted longer.
In all, results suggest that the lexical frequency of the governor modulates heritage speakers’ sensitivity to the lexical and structural conditioning of mood selection. Overall, high frequency governors boosted participants’ sensitivity to the fact that both Non-variable and negated governors favor the subjunctive (i.e., the difference between conditions was higher). In addition, smaller, but similar, effects are observed for lower frequency governors. Unlike Regularity, effects of frequency are observed both for Non-variable and negated governors. However, the magnitude of the effect was greater in Non-variable governors once again.

In the present study, it was also hypothesized that sociolinguistic variables might influence heritage speakers’ sensitivity to mood selection (Giancaspro, 2017; Gudmestad, 2006, 2010, 2012; Lubbers Quesada, 1998; Perez-Cortes, 2020). The results for the analyses of these factors are described below.
Sociolinguistic Variables

**Proficiency.** Proficiency was measured by participants’ scores in the Spanish block of the Picture Naming task they completed (see the Participants subsection); lower RTs when naming pictures in Spanish reflected greater proficiency in the language, indicating that participants could access individual lexical items faster.

- For the *Lexical Conditioning* comparison (see model output in Appendix A, Table G), Figure 3-14 shows a gradient effect: the higher the proficiency (i.e., lower RTs) the greater the difference in pupillary response between conditions. The strongest significant effect was found for high proficiency, starting approximately at bin 70 (140 ms into the target period), while a weaker significant effect for lower proficiency started approximately at bin 35 (70 ms into the target period). The difference smooths showed that both the magnitude and duration of the effect were greater for higher proficiency, although the effect for lower proficiency peaked earlier.

**Figure 3-14.** Difference in Pupillary Response based on Fitted Values
Proficiency – Non-variable Governors
• For the *Structural Conditioning* comparison, Figure 3-15 shows the same gradient pattern of results (see model output in Appendix A, Table H): the significant difference in pupillary response between conditions was stronger for higher proficiency, starting approximately at bin 100 (200 ms into the target period), and a weaker significant effect for lower proficiency started approximately at bin 40 (80 ms into the target period). The difference smooths showed that the magnitude of the effect was greater for higher proficiency, although the effect for lower proficiency peaked earlier and the duration was similar.

![Difference in Pupillary Response based on Fitted Values](image)

**Use of the Heritage Language.** When completing the LHQ for this study, participants used a percentage range slider to indicate how much they used the heritage language overall per week (M= 92.26%, SD= 17.66, range= 15).
• For the *Lexical Conditioning* comparison (see model output in Appendix A, Table M), Figure 3-16 below shows a gradient pattern, indicating that higher use of the heritage language lead to greater the difference in pupillary response between conditions. The strongest significant effect was found for high use, starting approximately at bin 45 (90 ms into the target period), while a weaker significant effect for lower use started approximately at bin 20 (40 ms into the target period) and was weaker. The difference smooths showed that both the magnitude and duration of the effect were greater for higher use of the heritage language, although the effect for lower use peaked earlier.

![Figure 3-16. Difference in Pupillary Response based on Fitted Values Heritage Language Use – Non-variable Governors](image)

• For the *Structural Conditioning* comparison, Figure 3-17 shows the same gradient pattern of results (see model output in Appendix A, Table J). The difference in pupillary response between conditions was significantly
greater when use of the heritage language was higher starting approximately at bin 100 (200 ms into the target period), and a weaker significant effect for lower use started approximately at bin 55 (110 ms into the target period). Again, the difference smooths showed that the magnitude and duration of the effect were greater for higher use, although the effect for lower use peaked earlier.

**Figure 3-17.** Difference in Pupillary Response based on Fitted Values Heritage Language Use – Negated Governors

---

**Language Entropy.** Entropy is a continuous measure that indexes the relative balance or diversity in the daily usage of two or more languages: higher entropy values relate to more balanced language use and greater language diversity (Gullifer & Titone, 2019: 4). Values range continuously from 0 to 1, in the current study participants’ mean entropy score was 0.75 (SD=.07, range=.05)\(^\text{15}\).

\(^{15}\) Entropy was calculated using the materials provided by Gullifer and Titone (2019); https://github.com/jasongullifer/languageEntropy
• For the *Lexical Conditioning* comparison (see model output in Appendix A, Table K), Figure 3-18 shows a gradient pattern: higher entropy scores lead to greater the difference in pupillary response between conditions. The strongest significant effect was found for high entropy, starting approximately at bin 90 (180 ms into the target period), while a weaker significant effect for lower entropy started approximately at bin 45 (90 ms into the target period). The difference smooths showed that the magnitude of the effect was greater for higher entropy, although the effect for lower entropy peaked earlier and lasted longer.

**Figure 3-18.** Difference in Pupillary Response based on Fitted Values
Language Entropy – Non-variable Governors

• For the *Structural Conditioning* comparison, Figure 3-19 shows a dichotomous pattern in results (see model output in Appendix A, Table L): the difference in pupillary response between conditions was
significantly greater for higher entropy scored starting approximately at bin 50 (100 ms into the target period) and so did a weaker significant effect for lower entropy use, although the effect was weaker. The difference smooths showed that the magnitude and duration of the effect were greater for higher entropy, although the effect for lower entropy peaked earlier.

**Figure 3-19.** Difference in Pupillary Response based on Fitted Values Language Entropy – Negated Governors

Identification with the Heritage Language. When completing the LHQ for this study, participants used 7-point Likert Scale (1 = “Very Low”, 7 = “Perfect”) to indicate how much identified with the heritage language (M= 6.18, SD= 1.13, range = 2).

- For the Lexical Conditioning comparison (see model output in Appendix A, Table M), Figure 3-20 below shows a gradient pattern of results indicating that higher identification with the heritage language lead to greater
difference in pupillary response between conditions. The strongest significant effect was found for high identification, starting approximately at bin 80 (160 ms into the target period), while a weaker significant effect for lower identification started approximately at bin 30 (60 ms into the target period). The difference smooths showed that both the magnitude and duration of the effect were greater for higher identification with the heritage language, although the effect for lower identification peaked earlier.

**Figure 3-20.** Difference in Pupillary Response based on Fitted Values Identification with the Heritage Language– Non-variable Governors

- For the *Structural Conditioning* comparison, **Figure 3-21** shows almost the same pattern of results (see model output in Appendix A, Table N): the difference in pupillary response between conditions was once again significantly greater for higher identification starting approximately at bin 70 (140 ms into the target period), while the significant effect for lower
identification started at bin 45 (90 ms into the target period). The difference smooths showed that the magnitude of the effect was greater for higher entropy, although the effect for lower entropy peaked earlier and duration was the same.

**Figure 3-21.** Difference in Pupillary Response based on Fitted Values
Identification with the Heritage Language– Negated Governors

The results of the analysis conducted with sociolinguistic variables suggest that they modulate heritage speakers’ sensitivity to the *Lexical* and *Structural Conditioning* of mood selection. Overall, higher the proficiency, use of the heritage language, language entropy and identification with the heritage language boosted participants’ sensitivity to the fact that both Non-variable and negated governors favor the subjunctive. In addition, smaller, but similar, effects are observed for lower frequency governors. Once again, these effects were non-selective and were observed both for Non-variable and Negated...

---

16 The weak significant effect starting at bin 0 is potentially spurious given that time is needed within the target period for the pupil to actually show a response to the stimulus.
governors. However, the magnitude of the effect was greater in Non-variable governors across the board.

To summarize the results presented so far, the heritage speakers in the current sample have shown to be sensitive to the Lexical and Structural Conditioning of mood selection in Spanish. Critically, this sensitivity is modulated by both linguistic and sociolinguistic factors. In addition, it was observed that the magnitude of the effects was greater for those conditions manipulating lexical factors (i.e., Lexical Conditioning).

**Experiment 2: Elicited Production**

**Participants.** Six months after participating in the auditory pupillometry experiment, the same 33 heritage speakers were contacted to take part in the elicited production experiment. A total of 20 individuals answered and participated in this task.

**Materials**

**Experimental comparisons and counterbalancing.** This experiment compared the production of subjunctive with Non-variable governors (Lexical Conditioning) and Negated governors (Structural Conditioning). It was predicted that participants would produce a subjunctive form when the main verb was a Non-variable or negated governor, both of which favor the subjunctive (Torres Cacoullos et al., 2017; Poplack et al., 2018). As in the previous experiment, all sentences were preceded by a context. The target region for analyses was a blank after the complementizer *que* (‘that’), which was to be completed with a subordinate verb of the participants’ choice. The following conditions were compared:
1. *Lexical Conditioning*: Condition *Non-variable governor* consisted of a Non-variable governor (in bold in Table 3-9) followed by a blank for the subordinate verb to be completed with the conjugated form of one of the three verbs provided.

<table>
<thead>
<tr>
<th>Original token</th>
<th>Contextual sentence</th>
<th>Condition</th>
<th>Target Sentence</th>
<th>Verb options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Le dijo que</td>
<td>Mi abuela tiene algunos problemas médicos en las piernas y por eso vamos con ella al hospital. Mi mamá y yo estamos hablando con el doctor sobre qué puede hacer para sentirse mejor..</td>
<td>Non-variable governor</td>
<td>Finalmente, el doctor le dice que ______ todos los días.</td>
<td>a. caminar</td>
</tr>
<tr>
<td>caminara</td>
<td></td>
<td></td>
<td></td>
<td>b. andar</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c. moverse</td>
</tr>
</tbody>
</table>

**Translation**

'S/he told him/her to walk

‘My grandma has some medical issues in her legs, so we took her to the hospital. My mom and I are talking with a doctor about what my grandma can do to feel better.”

‘Finally, the doctor tells her to ______ every day.’

a. to walk b. to stroll c. to move

2. *Structural Conditioning*: Condition *Negated governor* consisted of a negated governor (in bold in Table 3-10) followed by a blank for the subordinate verb to be completed with the conjugated form of one of the three verbs provided
Table 3-10. Elicited Production – Structural Conditioning
Experimental stimuli

<table>
<thead>
<tr>
<th>Original token</th>
<th>Contextual sentence</th>
<th>Condition</th>
<th>Target Sentence</th>
<th>Verb options</th>
</tr>
</thead>
</table>
| No creo que sea cierto eso, eres alérgico a los perros | Hace unos días, mi marido adoptó un perrito, pero ahora no puedo parar de estornudar. Yo quiero quedármelo, pero él piensa que soy alérgica. | Negated governor | Por eso, mi marido no cree que ______ bueną idea quedárnoslo. | a. parecer  
b. ser  
c. resultar |

Translation

‘I don’t think- gov, that is, sub, true, you are allergic to dogs’

‘Some days ago, my husband adopted a puppy, and I can’t stop sneezing since then. I want to keep it anyway, but my husband thinks I’m allergic to it’

‘For that reason, my husband doesn’t think gov that it ______ a good idea to keep it.’

a. to seem  
b. to be  
c. to appear

For the Lexical Conditioning and the Structural Conditioning comparisons, 50 sentences with Non-variable governors and 50 sentences with Negated governors were chosen. This resulted in a total of 100 experimental sentences. In addition, 25 filler sentences and 2 practice sentences were also included, resulting in a total of 127 items. Unlike the auditory pupillometry experiment, the elicited production task took place in a single session; the number of items was slightly reduced in order to decrease the time that it would take participants to complete the task and, thus, avoid possible adverse effects of fatigue on the data. Finally, to prevent order effects, two lists were created each of which contained all materials presented in a different randomized order. The flowchart below (Figure 3-22) shows the distribution of items per list:
**Design of materials.** The sentences used were a subset of the same materials in the auditory pupillometry experiment. For each sentence, the three verb forms provided to complete the subordinate verb blank were selected such that (1) they were semantically plausible options to fill the blank, and (2) they were synonymous with each other, to the extent that it was possible. For instance, as shown in Table 3-9, the options *caminar*, *andar* and *moverse* are all synonymous and plausible options to complete the blank in the sentence ‘Finalmente, el doctor le dice que *caminar*/andar/moverse todos los días.’

**Experimental design.** The recorded stimuli used were a subset of the same audiofiles used in the auditory pupillometry experiment. However, rather than being spliced, the audio for the target sentences was cut after the complementizer *que* (‘that’) to create a blank for the subordinate verb. Thus, while participants could still read the end of the sentence on the screen, the audio stopped at the complementizer, as shown in Table 3-11 below.

<table>
<thead>
<tr>
<th>Table 3-11. Cutting of experimental sentences – Elicited Production</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main</strong></td>
</tr>
<tr>
<td>Al final, su jefe autoriza que</td>
</tr>
<tr>
<td><em>Translation:</em> ‘Her boss authorizes that she ______ during the weekends.’</td>
</tr>
</tbody>
</table>
**Procedure.** Participants completed the experiment while on a Zoom call with the researcher, a Google sites web page was created to conduct the experiment online; this web page contained:

1. A form with information about the experiment and a request for verbal consent.
2. Zoom set up instructions to ensure that participants could communicate with the researcher properly during the experiment.
3. A practice version of the elicited production task.
4. An experimental version of the elicited production task.

The task was implemented using a Google slides presentation; when a slide appeared on the screen, the audio started playing automatically so participants could simultaneously listen and read the stimulus. Importantly, the task was self-paced to allow participants to advance to the following slide whenever they felt ready and to listen to and read the text as many times as necessary. Figure 3-23 shows the design of a trial: participants were first presented with a context, after which they proceeded to the target sentence. When the audio for the target sentence stopped at *que* (‘that’) participants chose one of the three verb forms provided and completed the blank out loud. For instance, a possible response to the target sentence in Figure 3-23 would have been: *que camine todos los días* (‘that s/he walks every day’).
Participants’ verbal responses were recorded during the Zoom call, and the researcher transcribed and coded their answers in real time. Participants were paid at the end the session at a rate of 25 USD per hour, and a short debriefing was held where they were informed of the general goals of the study and were asked about their overall thoughts on the various tasks.

**Data Analysis.** The dependent variable analyzed was *response accuracy*. Participants’ responses were coded as correct (1) if they produced a subjunctive form for the subordinate verb blank after a Negated or Non-variable governor; responses were coded as incorrect (0) if participants produced any verbal form that was not the subjunctive. In addition, the verb form produced participants’ responses were coded as “regular” or “irregular” following the same criteria that was used in the auditory pupillometry experiment.
Prior to fitting the models, the independent categorical variables condition and regularity were centered using -0.5/0.5 sum contrasts. This coding scheme compares the mean of the dependent variable for a given level of a factor to the overall mean of the dependent variable. This makes the intercept, the estimates of variables and any potential interactions more interpretable, and accounts for potential imbalances in the number of trials by condition. In addition, centering is critical in the study at hand given that interactions between categorical and continuous variables were modeled; the coefficients of such interactions can only be interpreted meaningfully if the categorical variables have been previously specified to be a contrast centered at 0 (Engqvist, 2005; Gelman & Hill, 2007; Schielzeth, 2010; Hamaker & Muthén, 2019).

Linear mixed-effects models were fitted using the buildmer function (buildmer package v. 1.6) in the statistical software R (v. 4.0.0, R Core Team, 2020). This function uses lmer from the lme4 package but allows for a systematic and replicable way of simplifying random effects structures and testing fixed effects. The function starts by attempting to fit the most maximal model possible. If the model fails to converge, the function then simplifies the random effects structure via backwards stepwise elimination; in other words, it attempts to find the maximal random effects structure that still allows the model to converge. Once the maximally converging model has been identified, the function calculates p-values for all fixed effects based on Satterthwaite denominator degrees of freedom using the lmerTest package (v. 3.1-2); the resulting models are the maximally converging models that the data were able to support (Bates, Kliegl, Vasishth & Baayen, 2015).
The maximal model submitted to buildmer included the two-level fixed effects of condition (NVS or Neg-S)\(^7\), regularity (irregular or regular) and frequency, plus all two- and three-way interactions between them. Random intercepts for participant and item were included, with random slopes maximally specifies for the three-way interaction between condition, regularity, and frequency.

**Results**

The maximally converging model selected by buildmer supported all the specifications initially included in the maximal model (see Data Analysis subsection). The output of this model is shown in Table 3-12 below.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>SE</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.78</td>
<td>0.07</td>
<td>59.03</td>
<td>10.72</td>
</tr>
<tr>
<td>Condition</td>
<td>-0.36</td>
<td>0.15</td>
<td>55.02</td>
<td>-2.46</td>
</tr>
<tr>
<td>Frequency</td>
<td>0.09</td>
<td>0.04</td>
<td>51.77</td>
<td>2.03</td>
</tr>
<tr>
<td>Regularity</td>
<td>0.25</td>
<td>0.09</td>
<td>58.03</td>
<td>2.65</td>
</tr>
<tr>
<td>Condition by Frequency</td>
<td>0.17</td>
<td>0.09</td>
<td>50.93</td>
<td>1.93</td>
</tr>
<tr>
<td>Condition by Regularity</td>
<td>0.53</td>
<td>0.19</td>
<td>59.62</td>
<td>2.75</td>
</tr>
<tr>
<td>Frequency by Regularity</td>
<td>-0.76</td>
<td>0.05</td>
<td>59.95</td>
<td>-8.06</td>
</tr>
<tr>
<td>Condition by Frequency by Regularity</td>
<td>-0.34</td>
<td>0.11</td>
<td>60.82</td>
<td>-3.19</td>
</tr>
</tbody>
</table>

The maximal model revealed main effects of condition, regularity and frequency, as well as significant two- and three-way interactions between these variables. Figure 3.24 visualizes the average accuracy across conditions: as can be readily seen, the mean accuracy of condition Non-variable governor is higher than that of condition Negated governor.

\(^7\) The reference level of each factor is underlined.
Figure 3-24. Difference in Subjunctive Production Accuracy by Condition

Figure 3-25 below illustrates the three-way interaction between condition, regularity, and frequency by plotting the average accuracy across these variables. For graphing purposes, the continuous variable frequency was transformed into a two-level categorical factor (High vs. Low Frequency). The graphs show that, overall, accuracy was higher for high frequency items and for irregular items. The interactions can be appreciated where accuracy was highest for high frequency irregular items (upper left quadrant) and lowest for low frequency regular items (bottom right quadrant).
To summarize, the results of the elicited production experiment go in line with those of the auditory pupillometry, showing that both morphological regularity and lexical frequency play a role not just in the comprehension but also in the production of the subjunctive by heritage speakers of Spanish. In addition, also in line with the pupillometry results discussed before, it was observed that the magnitude of effects was greater for condition Non-variable governor.

**Group 2: Spanish-dominant bilinguals**

**Experiment 1: Auditory Pupillometry Study**

**Participants.** A group of 18 Spanish-dominant bilingual native speakers of Mexican Spanish participated in this study (9 females, mean age 28.12 years old, SD=
6.22, range= 20). All participants were in the process of completing a degree at the University of New Mexico at the time of data collection and were recruited from graduate-level courses. Besides the sociolinguistic questionnaire, participants completed three behavioral tasks that were proxies for language proficiency: Verbal Fluency task, Bilingual Picture Naming task and Elicited Imitation Task. In the following subsections, a concise summary of participants’ demographics, language background and language use is provided to acquaint the reader with the Spanish-dominant sample. In addition, the results of each of the behavioral tasks are given to provide a nuanced picture of participants’ proficiency.

**Demographic and linguistic characteristics.** Participants completed the same questionnaire as the heritage speaker group. The demographic questions revealed that all participants had been born in Mexico, while the remaining seven were born in Mexico and had completed their Elementary and Secondary education in this country. Participants’ reported having resided in the US for an average of 6.26 years (SD= 5.51). All participants’ parents had also been born and raised in Mexico and lived there at the time. Regarding their socioeconomic background, all parents had at least a Bachelor’s degree, 3 parents had Masters’ degree, and only one had a Doctoral degree.

The questions eliciting linguistic information were designed to characterize participants’ life-long linguistic experience with Spanish and English; information was elicited about participants’ age of exposure to Spanish and English, perceived dominance in both languages, and language use in different contexts. In addition, participants were asked to use a 7-point scale (from 1-Very low to 7-Perfect) to self-rate their proficiency in reading, writing, speaking, and understanding each language; the self-rating was administered twice during the data collection procedure: first at the end of the first
session, and then at the end of the second session. This was done to roughly ascertain the relative reliability of the measure.

The questions collecting language use data showed that all participants reported using Spanish significantly more on a weekly basis ($t(50) = -34.10, p < .001$). They also reported having been exposed to Spanish significantly earlier than English ($t(20) = -12.10, p < .001$) and considered Spanish to be their dominant language. It follows that self-ratings revealed that participants felt more proficient in Spanish than in English on average. Participant characteristics are summarized in Tables 3-13 and 3-14.

**Table 3-13. Participants’ self-reported characteristics**

<table>
<thead>
<tr>
<th>Measure</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>28.12</td>
<td>6.22</td>
</tr>
<tr>
<td>Age of acquisition: Spanish</td>
<td>0.12</td>
<td>0.02</td>
</tr>
<tr>
<td>Age of acquisition: English</td>
<td>10.13</td>
<td>5.93</td>
</tr>
<tr>
<td>Spanish Use</td>
<td>75.45%</td>
<td>2.47</td>
</tr>
<tr>
<td>English Use</td>
<td>24.55%</td>
<td>18.47</td>
</tr>
</tbody>
</table>

**Table 3-14. Self-reported proficiency (7-point Likert scale)**

<table>
<thead>
<tr>
<th>Language</th>
<th>Skill</th>
<th>Session 1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Spanish</td>
<td>Speaking</td>
<td>6.65</td>
<td>0.70</td>
<td>6.79</td>
</tr>
<tr>
<td></td>
<td>Reading</td>
<td>6.65</td>
<td>0.79</td>
<td>6.66</td>
</tr>
<tr>
<td></td>
<td>Writing</td>
<td>6.47</td>
<td>0.80</td>
<td>6.55</td>
</tr>
<tr>
<td></td>
<td>Understanding</td>
<td>6.82</td>
<td>0.53</td>
<td>6.71</td>
</tr>
<tr>
<td>English</td>
<td>Speaking</td>
<td>5.63</td>
<td>0.81</td>
<td>5.79</td>
</tr>
<tr>
<td></td>
<td>Reading</td>
<td>5.81</td>
<td>0.98</td>
<td>5.72</td>
</tr>
<tr>
<td></td>
<td>Writing</td>
<td>5.88</td>
<td>1.26</td>
<td>5.35</td>
</tr>
<tr>
<td></td>
<td>Understanding</td>
<td>5.69</td>
<td>1.01</td>
<td>5.68</td>
</tr>
</tbody>
</table>
Participants were also asked to specify which language they used most frequently: at home, school, work, free time and when engaged in hobbies (they used a 7-point Likert scale from 1-Never to 7-Always); the purpose of this question was to gain a measure of participants’ diversity of language use across different social contexts. As can be seen in Table 3-15 below, use of Spanish is higher across all contexts except for Work and School given that English is the majority language and, therefore, used more frequently in formal and professional contexts.

<table>
<thead>
<tr>
<th>Context</th>
<th>Spanish M</th>
<th>Spanish SD</th>
<th>English M</th>
<th>English SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home (/7)</td>
<td>7.00</td>
<td>0.00</td>
<td>1.97</td>
<td>1.01</td>
</tr>
<tr>
<td>Work (/7)</td>
<td>2.26</td>
<td>1.24</td>
<td>6.17</td>
<td>1.40</td>
</tr>
<tr>
<td>School (/7)</td>
<td>3.66</td>
<td>2.01</td>
<td>4.94</td>
<td>1.66</td>
</tr>
<tr>
<td>Free time (/7)</td>
<td>5.69</td>
<td>1.49</td>
<td>3.89</td>
<td>1.67</td>
</tr>
<tr>
<td>Hobbies (/7)</td>
<td>5.54</td>
<td>1.61</td>
<td>3.06</td>
<td>1.66</td>
</tr>
</tbody>
</table>

**Behavioral tasks**

**Verbal fluency.** Results showed that participants named significantly more exemplars in Spanish (M= 47.15) than in English (M= 39.25; t(61)= 4.23, p = 0.04).

**Picture naming task.** Results showed that, overall, participants were significantly more accurate in Spanish (M= 74.13%) than in English (M= 56.13%; t(60) = 7.75, p < 0.03), and significantly faster to respond in Spanish (M= 1149.47 ms) than in English (M= 1213.07 ms; t(50) = -1.14, p = 0.04).

**Elicited imitation task.** Results showed participants performed at ceiling, scoring 143.31 on average (SD= 0.97, range= 141-144).

**Results**
**Lexical Conditioning.** Figure 3-26 below shows the fitted smooths for the NVS and NVI conditions extracted from the main model of the *Lexical Conditioning* comparison. Descriptively, condition *Non-variable Indicative* elicited a larger pupillary response (higher magnitude and duration) starting at bin 35 (7000 ms into the target period) in comparison to condition *Non-variable Subjunctive*:

![Figure 3-26. Pupillary Response Smooths based on Fitted Values. Lexical Conditioning – Non-variable Governors](image)

The output of the model (Table 3-16) shows that the non-linear pattern for the difference between the two conditions (**in bold**) was significant:
Table 3-16. Main model output – Lexical Conditioning

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVS)</td>
<td>-40.51</td>
<td>4.33</td>
<td>-9.35</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Smooth terms

<table>
<thead>
<tr>
<th></th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVS)</td>
<td>9.86</td>
<td>12.26</td>
<td>2.10</td>
<td>.01</td>
</tr>
<tr>
<td>Diff. NVI minus NVS</td>
<td>5.44</td>
<td>6.77</td>
<td>2.22</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Gaze x by Gaze y</td>
<td>25.91</td>
<td>28.21</td>
<td>317.97</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>97.59</td>
<td>150.00</td>
<td>4.48</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>219.46</td>
<td>620.00</td>
<td>1.06</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .11; Deviance Explained: 11.20%

Structural Conditioning. Figure 3.27 below shows the fitted smooths for the Neg-S and Neg-I conditions extracted from the main model of the Structural Conditioning comparison. Descriptively, condition Negated Indicative elicited a larger pupillary response starting at bin 40 (800 ms into the target period) in comparison to condition Negated Subjunctive:
The output of the model (Table 3-17) shows that the non-linear pattern for the difference between the two conditions (in bold) was highly significant:

<table>
<thead>
<tr>
<th>Table 3-17. Main model output – Structural Conditioning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parametric coefficients</strong></td>
</tr>
<tr>
<td>(Intercept NVS)</td>
</tr>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>-30.66</td>
</tr>
<tr>
<td><strong>Smooth terms</strong></td>
</tr>
<tr>
<td>(Intercept NVS)</td>
</tr>
<tr>
<td>edf</td>
</tr>
<tr>
<td>5.57</td>
</tr>
<tr>
<td><strong>Diff. Neg-I minus Neg-S</strong></td>
</tr>
<tr>
<td>2.00</td>
</tr>
<tr>
<td>Gaze x by Gaze y</td>
</tr>
<tr>
<td>24.90</td>
</tr>
<tr>
<td>Time by Participant</td>
</tr>
<tr>
<td>72.87</td>
</tr>
<tr>
<td>Time by Item</td>
</tr>
<tr>
<td>269.12</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .15; Deviance Explained: 12.70%

**Linguistic Variables**

**Morphological Regularity.** Figure 3-28 below shows the fitted smooths for the Non-variable Subjunctive Regular/Irregular and Non-variable Indicative Regular/Irregular conditions. Descriptively, condition Non-variable Indicative Regular elicited a larger pupillary response from the very beginning of the target period in comparison to condition Non-variable Indicative Irregular. Likewise, condition Non-variable Subjunctive Regular elicited a larger pupillary response starting at bin 10 (200 ms into the target period) in comparison to condition Non-variable Subjunctive Irregular. As can be observed, the magnitude and duration of the difference in pupil-size is higher for the Non-variable Indicative conditions.
The output of the NVS model (Appendix A, Table O) showed that the non-linear pattern for the difference between the regular and irregular subordinate verbs was not significant. However, there was a significant interaction between condition and regularity \((F = 3.32, p < .001)\). The NVI model (Appendix A, Table P) showed both a significant effect of regularity (Regular: \(F = 3.23, p = .001\)) and a significant interaction \((F = 4.89, p < .001)\). Visualization of the fitted smooths (Figure 3-28) suggests that this interaction was largely driven by the significant difference between conditions Non-variable Subjunctive Regular and Non-variable Indicative Regular (in bold in Table P), where the effect of regularity was stronger for the Non-variable Indicative condition.

Fitted smooths for the Negated Subjunctive Regular/Irregular and Negated Indicative Regular/Irregular conditions are visualized in Figure 3-29. Descriptively, condition Negated Indicative Irregular elicited a larger pupillary response starting at bin 55 (1100 ms into the target period) in comparison to condition Non-variable Indicative Irregular. In addition, condition Negated Subjunctive Regular consistently elicited a larger pupillary response throughout the target period in comparison to condition Negated Subjunctive Irregular. That is, the magnitude and duration of the difference in pupil-size is higher for the Negated subjunctive conditions in this case.
For model Neg-S (Appendix A, Table Q), the only significant effect captured in was for the difference between conditions *Negated Indicative Irregular* and *Negated Subjunctive Irregular* ($F = 2.86, p = .05$) where the pupillary response was higher for the indicative condition. For the Neg-I model (Appendix A, Table R), however, the non-linear pattern for the difference between the regular and irregular subordinate verbs was not significant, but the model showed a significant difference between conditions *Negated Indicative Regular* and *Negated Subjunctive Regular* ($F = 3.04, p = .01$), as well as an interaction between condition and regularity ($F = 3.19, p < .001$).

To sum up, the results of the analyses conducted indicate that morphological regularity does modulate Spanish-dominant bilinguals’ sensitivity to mood selection. Overall, regular subordinate verbs are harder to process than irregular subordinate verbs; this difficulty is exacerbated when the governor requires the subjunctive, but the subordinate verb is in the indicative.
**Lexical Frequency**

For the *Lexical Conditioning* comparison (see model output in Appendix A, Table S), Figure 3-30 below shows a gradient significant difference in pupillary response between conditions for high frequency governors starting from the beginning of the target period. The difference smooths showed that both the magnitude and duration of the frequency effect were greater for higher frequency governors, which also peaked earlier.

**Figure 3-30.** Difference in Pupillary Response based on Fitted Values
Frequency – Non-variable Governors

For the *Structural Conditioning* comparison (see model output in Appendix A, Table T), Figure 3-31 shows there was a weaker but significant difference in pupillary response between conditions for both high frequency governors and low frequency governors, again from the beginning of the target period. The difference smooths showed that, once again, the magnitude of the frequency effect was greater for higher frequency governors, peaking earlier and lasting longer.
In all, results suggest that the lexical frequency of the governor modulates Spanish-dominant bilinguals’ sensitivity to the lexical and structural conditioning of mood selection. Overall, high frequency governors boosted participants’ sensitivity to the fact that both Non-variable and negated governors favor the subjunctive (i.e., the difference between conditions was higher, although the magnitude of the effect was greater in Non-variable governors.

**Elicited Production Experiment**

There were main effects of condition, regularity and frequency, as well as significant two-way interactions between these variables (Table 3-18).
Table 3-18. Elicited Production Model – Summary of Fixed Effects

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.99</td>
<td>0.01</td>
<td>158</td>
<td>0.99</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Condition</td>
<td>-0.04</td>
<td>0.01</td>
<td>-3.81</td>
<td>-0.04</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Frequency</td>
<td>-0.02</td>
<td>0.01</td>
<td>-2.87</td>
<td>-0.02</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Regularity</td>
<td>-0.02</td>
<td>0.01</td>
<td>-2.49</td>
<td>-0.02</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Condition by Frequency</td>
<td>0.27</td>
<td>0.30</td>
<td>70.63</td>
<td>3.53</td>
<td>.01</td>
</tr>
<tr>
<td>Condition by Regularity</td>
<td>0.23</td>
<td>0.11</td>
<td>99.13</td>
<td>4.92</td>
<td>.01</td>
</tr>
</tbody>
</table>

Figure 3-32, visualizes the average accuracy across conditions: as can be readily seen, the mean accuracy of condition Non-variable governor is higher than that of condition Negated governor.

Figure 3.32. Difference in Subjunctive Production Accuracy by Condition

To summarize, the results of the elicited production experiment go in line with those of the auditory pupillometry, showing that both morphological regularity and lexical frequency play a role not just in the comprehension but also in the production of the subjunctive by Spanish-dominant bilinguals. In addition, also in line with the
pupillometry results discussed before, it was observed that the magnitude of effects was
greater for condition Non-variable governor.

Discussion
The goal of the present study was to examine knowledge of the Spanish subjunctive in
heritage speakers who live in a long-standing bilingual community in Albuquerque, New
Mexico, in comparison to a sample of Spanish-dominant bilinguals from Mexico. To do
so, this study integrated corpus data into the design of psycholinguistic experiments that
tested participants’ sensitivity to the Lexical and Structural Conditioning of mood selection
in Spanish both in comprehension and production. In addition, to better understand the
nature of this sensitivity, it investigated several linguistic and sociolinguistic factors that
might modulate it. The results of this study showed that both in comprehension and
production: (1) the current sample of heritage speakers was highly sensitive to the lexical
and structural conditioning of mood selection, (2) both linguistic and sociolinguistic
factors modulated their sensitivity, and (3) the performance of heritage speakers and
Spanish-dominant bilinguals converged on the same results and trends.

While an ample body of literature has provided evidence that heritage speakers of
Spanish oftentimes fail to reach native-like knowledge of mood, a series of recent studies
have shown that heritage speakers can actually display robust and systematic knowledge
of the subjunctive (Bookhammer, 2013; Giancaspro, 2017; LaCasse, 2018; Perez-Cortes,
2016). The present study supports these recent findings and extends them by examining
a long-standing heritage speaker community and providing proof of concept of the
applicability of a new methodology to the field of heritage language research. In the
following subsections, I discuss the results of this study in more depth and consider the
findings in light of their relation and contribution to previous research.
Sensitivity to the Lexical vs. Structural Conditioning of mood selection

Results indicated that the heritage speakers in the current sample were highly sensitive to both the lexical and structural factors that condition mood selection in Spanish during comprehension and production. A pivotal finding was that the magnitude of the effect was smaller for the Structural Conditioning comparison, which was mirrored in the Spanish-dominant group. In the auditory pupillometry experiment, the difference between conditions (NVI minus NVS, and Neg-I minus Neg-S) was smaller for sentences within structurally conditioned contexts. Likewise, in the elicited production experiment, accuracy was overall lower for sentences within structurally conditioned contexts.

As explained in Chapter 1, subsection Acquisition of the subjunctive, the general consensus among scholars is that acquiring the subjunctive is no easy task, neither in the L1 nor in the L2:

- First, where verbs are not irregular (voy-1st sg. Ind. vs. vaya-1st sg. Subj., ‘to go’), subjunctive morphology is oftentimes not salient given that the only marker of mood is a final vowel change (quiero-1st sg. Ind. vs. quiera-1st sg. Subj., ‘to want’). This reduces the saliency of the subjunctive in the input (Collentine, 2003; Gudmestad, 2008).

- Second, the subjunctive’s syntactic distribution is largely limited to subordinate clauses which diminishes its frequency in the input (Biber et al., 2006).

- Third, the range of connotations of the subjunctive is opaque and nuanced, making it difficult for speakers to identify the subjunctive’s contribution to the meaning of utterances (Collentine, 2003).
These characteristics cause the subjunctive to have a protracted developmental schedule: children do not have full, adult-like mastery of all the nuances and uses of the subjunctive until after the age of 8 (Blake, 1983; Pérez-Leroux, 1998; Sánchez-Naranjo & Pérez-Leroux, 2006). Research on the acquisition of the subjunctive in monolingual and bilingual populations has shown that it is first learned in “obligatory” contexts of use, (i.e., intensional subjunctive, deontic predicates, Non-variable governors) and then in “optional” contexts of use (i.e., polarity subjunctive, epistemic/epistemological predicates, variable governors; Anderson, 1999, 2001; Blake, 1983; Hernandez Pina, 1984; Merino, 1983; Naharro, 1996; Lozano, 1974; Papafragou, 1998; 2010; Perez-Cortes, 2016; Silva-Corvalán, 2014). It has been hypothesized that maturational constraints mediate the acquisition of the subjunctive in different contexts (de Villiers & de Villiers, 2010; Perez-Leroux, 1998); for instance, Perez-Leroux (1998) argued that mood selection is first acquired with deontic predicates because these only necessitate comprehending meanings such as obligation and permission; by comparison, the implicatures of epistemic/epistemological predicates are more difficult to conceptualize because they require the consideration of the speaker’s point of view and their comments of the validity of the statement. Based on this, Perez-Leroux proposed that the understanding of false beliefs (irrealis) is a cognitive prerequisite for the acquisition of the evaluative values of the subjunctive.

Numerous empirical studies have shown that both monolinguals and bilinguals are more accurate in their interpretation and use of mood in obligatory contexts of use than in optional contexts of use (Ahearn et al., 2014; Blake, 19983; Echeverría, 1978; Gallo-Valdivieso, 1994; Geeslin, 2014; Iverson et al., 2008; Kanwitt &; Montrul, 2007, 2009; Pascual y Cabo et al., 2012; Padilla, 1990; Perez-Cortes, 2016). The present study also converged with these findings: the Non-variable governors included in the Lexical
Conditioning comparison were all volitional and desiderative verbs, which could be considered deontic predicates and, as such, obligatory contexts of subjunctive use; while the negated governors in the Structural Conditioning were all evaluative verbs, which are epistemic/epistemological predicates and belong to the category of optional contexts. In line with previous research, results indicated that the heritage speakers in the current sample were more sensitive to the lexical conditioning of mood selection than to its structural conditioning.

While this is a valid interpretation of the results observed, the characterization of mood selection in which this study is grounded is based on corpus data which have shown that the lexical identity of the governor is the primordial factor contributing to subjunctive selection in Spanish, independently of semantic, pragmatic and structural factors (see Chapter 1, subsection Variationist characterizations of the subjunctive). Thus, I propose an alternative explanation of the results of the present study based on the distribution of the factors conditioning mood selection in speakers’ input. The data analyzed by Torres Cacoullos et al., (2017) indicated that Non-variable governors are the most frequent lexical type in Spanish: 61% of the subjunctive sentences extracted by Torres Cacoullos et al., (2017) contained a Non-variable governor. By comparison, variable governors, that is, those where structural factors play a role in conditioning subjunctive selection, constituted only a third of the data (39%).

Considering the distribution of these conditioning factors in the input heritage speakers receive is essential for understanding how they acquire and process mood. Mood selection is a complex instance of syntactic alternation (i.e., subjunctive vs. indicative) whose variation is conditioned by several factors (Shin, 2016), thus making it more challenging to acquire. As children become attuned to the distributional patterns of mood alternation, they first acquire those conditioning factors which are more reliably
associated with the subjunctive; these are the most frequent ones, as they are easier to detect. Given that the subjunctive is primarily lexically conditioned and Non-variable governors are more frequent, it follows that the lexical conditioning of mood selection is acquired earlier. This notion is supported by studies on children’s acquisition of morphosyntactic variation which have shown that the sensitivity to conditioning contexts of linguistic alternation is first apparent in high-frequency contexts and high frequency lexical items (Shin, 2016). It has been suggested that variation is at first tied to frequent lexical items and then generalizes to infrequent items (Chevrot, Beaud, & Varga, 2000).

In light of the evidence discussed so far, the asymmetry between the lexical and structural conditioning of mood selection might be explained by the fact that Non-variable governors (i.e., volitional/desiderative verbs) are more frequent and, therefore acquired earlier, which strengthens their representations in heritage speakers’ grammars, making them more entrenched. By virtue of this, heritage speakers are more sensitive to lexical conditioning factors during online comprehension and are also more accurate in their production of the subjunctive in lexically conditioned contexts.

In the following subsection, I discuss the effects of regularity and lexical frequency, elaborating on why differences on how these factors modulate the lexical and structural conditioning of mood selection might be observed.

The effect of linguistic variables

While it has been established that mood selection in Spanish is primarily lexically conditioned, with structural factors playing a secondary role, there is evidence that other linguistic factors might also modulate speakers’ knowledge of the subjunctive. In this study, it was examined whether the morphological regularity of the subordinate verb and
the lexical frequency of the governor modulate heritage speakers’ sensitivity to the Lexical and Structural Conditioning of mood selection.

Morphological regularity, which refers to the regularity of the verb stem (e.g., voy/venga vs. quiero/quiera; ‘to go’, ‘to want’), has been shown to play a role in both monolinguals’ and bilinguals’ comprehension and production of the Spanish subjunctive. For instance, Gudmestad (2010, 2006) found that both monolingual speakers and L2 speakers of Spanish produced significantly more subjunctive forms with irregular subordinate verbs than regular subordinate verbs. In addition, it was also observed that irregular verbs best predicted subjunctive selection when preceded by volitional governors. Likewise, the effect of lexical frequency has been amply documented both in the L1 and the L2 (see Ellis, 2002 for a review). Frequency has been reported to modulate lexical access such that more frequent words are faster to access both for comprehension and production (Ellis & Collins, 2009; Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Wulff, Ellis, Römer, Bardovi-Harlig & Leblanc, 2009). Importantly, frequency has also been shown to influence the rate of acquisition of morphosyntactic structures, where more frequent structures are acquired earlier (Bybee & Hopper, 2001; Lieven & Tomasello, 2008; Mueller Gathercole, 2007; Paradis, 2010; Shin, 2016).

In comparison to other fields of research, morphological regularity and lexical frequency have been relatively understudied within in the context of heritage grammars. To my knowledge only two studies have examined how these factors modulate heritage speakers’ processing of the Spanish subjunctive: Giancaspro (2017) found that the regularity and frequency of the governor, and the subordinate verb, modulated heritage speakers’ accuracy in the production and interpretation of the subjunctive in several offline tasks, such that participants were more accurate with more frequent items. Similarly, Perez-Cortes (2020) found that that frequent governors and irregular
subordinate verbs yielded higher rates of accuracy in heritage speakers’ production and interpretation of the subjunctive in a picture-based sentence completion task.

In the present study, similar effects of regularity and frequency were found both in comprehension and production for heritage speakers. In the auditory pupillometry experiment, the effect of regularity was more robust for the *Lexical Conditioning* comparison, showing an interaction between regularity and condition which was driven by a significantly larger difference in pupillary response between conditions *Non-variable Indicative Regular* and *Non-variable Subjunctive Regular*. This indicates that during online comprehension, it was more costly for participants to process the mood alternations within lexically conditioned contexts when the subordinate verb was morphologically regular.

Likewise, there was an effect of frequency both for the *Lexical* and *Structural Conditioning* comparisons, albeit the magnitude was smaller for the latter. Overall, high frequency governors boosted participants’ sensitivity to the fact that both Non-variable and negated governors favor the subjunctive. Importantly, parallel results were reported for the elicited production experiment, where heritage speakers were more accurate in their production of the subjunctive when the subordinate verb was irregular and when the governor was frequent. Likewise, these effects were stronger for sentences within lexically conditioned contexts.

Given the extensive literature on the effects of lexical frequency, it is no surprise that this variable has a facilitatory effect on heritage speakers’ processing of the subjunctive for both comprehension and production. Lexical items that are more frequent in the input are more entrenched in speakers’ grammars and are, therefore, easier to access. Regarding regularity, Collentine (1997) hypothesized that irregular forms are more salient to speakers, which facilitates their interpretation and retrieval. While this
statement goes in line with previous research on the effects of saliency on accessibility (Burkhardt & Roehm, 2007; Khezrlou, 2020), it must also be noted that it is hard to disentangle frequency and regularity, given that morphologically irregular items are oftentimes high-frequency items (Bybee & Hopper, 2001).

In the present study, irregular governors were more frequent\textsuperscript{18} in the *Lexical Conditioning* comparison (regular: $M=1.99$, SD=.70; irregular: $M=2.29$, SD=.33), while regular governors were more frequent in the *Structural Conditioning* comparison (regular: $M=1.59$, SD=.38; irregular: $M=1.46$, SD=.37). This might explain, at least in part, why an effect of regularity was not observed for the *Structural Conditioning* comparison in the pupillometry experiment, and why the magnitude of the regularity effect was smaller for sentences within structurally conditioned contexts in the elicited production task. It is also possible that the magnitude of the frequency and regularity effects on the *Structural Conditioning* comparison is reduced in comparison to the *Lexical Conditioning* comparison because, as explained in the previous subsection, variable governors in which structural factors play a role in conditioning subjunctive selection are acquired later, as they are less frequent in the input, and are overall less entrenched in speakers’ grammars.

In all, while the magnitude of the effect differs depending on the conditioning context, the results discussed in this subsection provide evidence that heritage speakers possess robust and nuanced knowledge of the morphosyntactic factors that modulate the conditioning of mood selection.

\textsuperscript{18} These are log frequencies per million extracted from the LexEsp corpus (Sebastián-Gallés, 2000).
The effect of sociolinguistic variables

Proficiency is considered a pivotal factor in understanding heritage speakers’ linguistic knowledge (Kagan & Dillon, 2004; Kagan & Friedman, 2003; Kagan, 2005; Kupisch, 2013; Shea, 2017; Montrul, 2007, 2009; Swender et al., 2014) given that significant differences in receptive and productive use of the heritage language have been found among speakers at different levels in the proficiency spectrum (for a review see Benmamoun et al., 2013). With respect to mood selection, research on the acquisition of Spanish subjunctive has shown that heritage speakers with higher levels of proficiency discriminate more accurately between indicative/subjunctive interpretations and tend not to overextend indicative forms in production (Iverson et al., 2008; Dekydtspotter & Renaud, 2014; Mikulski, 2006; Montrul, 2007, 2009; Silva-Corvalán, 1994).

However, accurately assessing proficiency in heritage populations is particularly challenging; factors such as task modality, implicitness and register greatly impact heritage speakers’ performance in assessment tasks. Ample evidence has been provided that given their acquisitional background, heritage speakers tend to underperform in tasks and tests that are written, require the use of explicit knowledge and metalinguistic and skills, and those that employ formally standardized forms of the majority language (Menken & Kleyn, 2010; Montrul, Foote & Perpiñán, 2008; Otheguy, 2013). Achieving a theory of heritage language acquisition and processing that has greater explanatory adequacy (Carreira, 2004) requires that scholars look beyond proficiency. While there is evidence that socially-situated linguistic experience impacts language outcomes in monolinguals and bilinguals (Beatty-Martínez & Dussias, 2017; Dąbrowska, 2009, 2012; Dąbrowska & Street, 2006; Gudmestad, 2010, 2012, 2013; Guzzardo Tamargo et al., 2016; Johns et al., 2019; Valdés Kroff, 2012; Street & Dąbrowska, 2010), experience-based
variables are rarely included as explanatory factors in studies on bilingualism, especially within the field of heritage language studies.

In the case of the Spanish subjunctive in particular, sociolinguistic variables remain largely understudied (Gudmestad, 2010; Puente-Schubeck, 1991; Waltermire, 2014). The heritage speakers in the current sample completed a detailed sociolinguistic questionnaire designed to operationalize their linguistic experience along several dimensions (see Chapter 2); based on the PCA conducted on these data, the role of three sociolinguistic variables was examined in the present study: use of the heritage language, language entropy and identification with the heritage language. It was predicted that these sociolinguistic variables would modulate sensitivity to mood selection during online comprehension in the current sample of heritage speakers. Results indicated that this prediction was met for all variables, such that sensitivity to the Lexical and Structural Conditioning of mood selection was greater for heritage speakers who: (1) used the heritage language more on average, (2) used the heritage language in more diverse contexts, and (3) felt more identified with the heritage language. In addition, to use a proficiency assessment method that was as natural as possible in relation to participants’ experience with the heritage language, proficiency was measured using a Bilingual Picture Naming task (i.e., an online, oral and implicit task). The findings showed that heritage speakers with faster RTs in the BPN task in Spanish, an indicator that they were more proficient in the heritage language, showed greater sensitivity to the conditioning of mood selection.

In all, the results of the present study highlight the fact using appropriate assessments of proficiency and taking into consideration factors beyond proficiency are crucial to achieve a better understanding of bilinguals’ linguistic knowledge. In the case of heritage speakers, the high variability this population (for reviews see: Aalberse &
Muysken, 2013; Rothman & Treffes-Daller, 2014), makes it even more critical to have more studies that explore how sociolinguistic factors such as psychosocial and biographical variables modulate heritage grammars.

**The effect of the community of practice**

Asymmetries between comprehension and production abilities have oftentimes been observed in heritage speakers (Mikulski, 2006; Mikulski & Elola, 2013; Montrul, 2009, 2011; Perez-Cortes, 2016; Rothman, Pascual y Cabo & Lingwall, 2012; Sherkina-Lieber, Pérez-Leroux & Johns, 2011; Valdes Kroff, Villegas and Dussias, under review). Putnam and Sánchez (2013) and Perez-Cortes et al. (2019) have proposed that the differential outcomes observed in heritage speakers’ production, but not so much comprehension, are the result of an early shift in language dominance due to which the majority language (L2) becomes dominant, more activated, and easier to access than the heritage language (L1). Differential activation of competing representations for production and comprehension in both languages may eventually lead to complications accessing these representations in the non-dominant language, especially in production. In the case of mood, they argue that elements of grammar that are less salient (i.e., opaque form and meaning, los frequency) are especially impacted by the continued activation of the dominant L2 throughout the remainder of the lifespan of heritage speakers.

In the present study, asymmetries between heritage speakers’ sensitivity to the Lexical and Structural Conditioning of mood selection in comprehension and production were not observed. In fact, participants displayed robust, systematic, and nuanced knowledge of the morphosyntactic intricacies of mood in Spanish. While the design of the materials and the methodology employed might indeed have played a role, it is also possible that the results observed here diverge from those of previous studies due to the
population sampled. As explained in Chapter 1, New Mexico is home to the oldest Spanish speaking community in the US and, nowadays, it has the highest Mexican origin population in the US (46.7% of the state speaks Spanish). Thus, New Mexico has been a highly bilingual community for centuries (Travis & Torres Cacoullos, 2013; Villa 2003; Mora, Villa & Dávila, 2005). While there is a dearth of research investigating long-standing heritage communities, there are a few studies that have examined heritage speakers in Albuquerque, New Mexico before. The most notable work characterizing this population is that of Travis and Torres Cacoullos (2018) who compiled the New Mexico Spanish-English Bilingual (NMSEB) corpus; the informants in the NMSEB had to fit the following criteria:

- be no less than third generation Nuevomexicanos (‘New Mexicans’).
- regularly use both languages with the same interlocutor in the same contexts.
- have high proficiency self-ratings in each language.

Data from the NSMEB shows that the population sample belongs to a highly bilingual community who often combine their two languages in discourse, engaging in dense code-switching. LaCasse (2018) examined production of subjunctive in heritage speakers within the NMSEB, using the Corpus Sociolingüístico de la Ciudad de México (CSCM; Martín Butragueño, 2011, 2012, 2015) as a benchmark for language change to investigate whether language contact with English was causing a progressive reduction of subjunctive forms in New Mexican Spanish. Her results showed no evidence of

---

19 see Torres Cacoullos and Travis (2018) for more information about the speech community and the careful compilation of the corpus.
simplification in the grammars of New Mexican heritage speakers, who used the subjunctive with comparable frequency and under the same linguistic conditioning as the monolingual benchmark. In all, the results reported by LaCasse help better understand why an asymmetry between production and comprehension of the Spanish subjunctive was not found in the present study (for similar results see Johns & Steuck, 2021). These findings speak to the relevance of the linguistic community for heritage populations, bringing into focus its crucial role in shaping the language knowledge of its members through experience with socially-situated language acquisition and use (Eckert, 2006; Gumpertz & Hymes, 1972; Labov, 1972).

Conclusion
The experiments in this chapter have the potential to advance the field of heritage language research by providing a more holistic understanding of heritage grammars, thus contributing to a theory of language acquisition and processing in heritage language contexts that has greater explanatory adequacy. The findings reported highlight that factors such as the community examined, the ecological validity of the materials used and the diversity of explanatory variables included in analyses are crucial. In addition, they underscore the importance of triangulating both comprehension and production experimental data for a more comprehensive approach to complex and highly variable systems such as heritage grammars.
CHAPTER 4: DISCUSSION

The goal of the present dissertation was to achieve a holistic understanding of the nature of grammars of heritage speakers, and to contribute to a theory of processing in heritage language contexts that has greater explanatory adequacy. To this end, this dissertation set out to provide (1) an evidence-based characterization of heritage speakers by using a sociolinguistic questionnaire, which along with PCA, examined the language-experience related factors that best explain the variability in the processing of the subjunctive mood in this population; and (2) an examination of heritage speakers’ and Spanish-dominant bilinguals’ processing of the Spanish subjunctive during online comprehension and production by means of psycholinguistic experiments that integrated corpus data into their design.

In the following subsections, I will first provide a brief overview of the results reported in this dissertation. In light of these results, I will then follow with a discussion of why a prescriptive treatment of subjunctive mood selection hinders progress in the field of heritage language research and will discuss how a usage-based approach to language can help scholars achieve a more adequate and unified theory of heritage grammars. In ending, I discuss several future directions.

Main Findings

A. Principal Component Analysis: Empirically characterizing variability in heritage communities

In this chapter, it was argued that discrepancies in the results reported by studies with heritage speakers might, at least in part, be accounted for by the way heritage populations are characterized. To address this issue, bilingual experience was operationalized into
four core dimensions, and Principal Component Analysis was used as an empirical method to explore these dimensions to maximize the explanatory power of LHQ data. The results for the first dimension, *Heritage Language Background*, underscore the relevance of early linguistic experience and its interaction with demographics in the shaping of heritage speakers’ linguistic knowledge. The findings also highlight the fact that the linguistic background of heritage speakers crucially affects their self-rated judgments, which can make this measure unreliable to study this population. For the second dimension, *Contexts of Use*, results indicated that the relative richness of contexts of use (*Free time, Hobbies* and *School*) was the factor that explained most of the variance in the data. This suggests that variability in the contexts of use of the heritage language can be used as a fine-grained measure of differentiation of individuals within heritage populations. The third dimension explored the productivity and bilingualism of participants’ *Interactional Networks* (immediate family, extended family, friends and acquaintances). Results suggested that it was the productivity and bilingualism of the immediate and extended family networks that best differentiated participants within the current sample of heritage speakers. Given that early experience is primarily provided by family members, having a large, productive and highly bilingual family network is crucial to promote bilingualism. Finally, results for the fourth dimension, *Attitude, Motivation and Identity*, showed that the psychosocial variables that explained the most variance in the data were identification with the heritage language and perception of the heritage language. Thus, the examination of the data from the sociolinguistic questionnaire in light of the results of the Principal Component Analysis brings attention to the relevance of affective factors, showing their potential for the characterization of heritage populations.
In all, the descriptions and differentiations of the different heritage populations will directly impact the interpretation and extrapolation of results across studies, as well as the theories derived from them. Principal Component Analysis is an empirical method that can help scholars generate evidence-based characterizations of heritage speakers by examining which language-experience-related factors best explain the variability inherent to a sample.

B. Comprehension and Production Results: Convergence not divergence

Studies assessing heritage speakers of Spanish grammatical knowledge of the subjunctive mood have concluded, almost unanimously, that heritage speakers’ knowledge of the Spanish subjunctive is non-native-like. However, a series of recent studies have challenged these deficit-oriented perspectives providing evidence that, while different from that of non-heritage speakers, heritage speakers’ linguistic knowledge is by no means deficient. In line with this research, the current study compared the performance of heritage speakers of Spanish in a comprehension (auditory pupillometry) and a production (elicited production) tasks to that of a sample of Spanish-dominant bilinguals. The following research questions were addressed:

Research question 1: *Are heritage speakers who live in a long-standing bilingual community sensitive to both the lexical and structural factors that condition mood selection in Spanish?*

Results indicated that the current sample of heritage speakers was highly sensitive to the lexical and structural conditioning of mood selection, displaying robust and systematic knowledge of the subjunctive. The magnitude of the effect was smaller for the structurally conditioned contexts, an effect which was mirrored in the Spanish-dominant group.
Research question 2: *What other linguistic aspects modulate the degree of sensitivity displayed by heritage speakers?*

While it has been established that mood selection in Spanish is primarily lexically conditioned, with structural factors playing a secondary role, there is evidence that other linguistic factors might also modulate speakers’ knowledge of the subjunctive. In this dissertation, I examined whether the morphological regularity of the subordinate verb and the lexical frequency of the governor modulated heritage speakers’ sensitivity to the *Lexical* and *Structural Conditioning* of mood selection. Results indicated that it was more costly for heritage speakers to process mood alternations in contexts where the subordinate verb was morphologically regular. In addition, high frequency governors boosted heritage speakers’ sensitivity to the fact that both Non-variable and negated governors favor the subjunctive. Importantly, the magnitude of the regularity and frequency effects was greater in lexically conditioned contexts for both heritage speakers and Spanish-dominant bilinguals.

Research question 3: *Do sociolinguistic factors also modulate heritage speakers’ sensitivity to mood selection?*

The heritage speakers in the current study completed a detailed sociolinguistic questionnaire designed to operationalize their linguistic experience along several dimensions; based on the PCA conducted on these data, the role of three sociolinguistic variables was examined: *use of the heritage language, language entropy* and *identification with the heritage language*. As predicted, results indicated that sensitivity to the *Lexical* and *Structural Conditioning* of mood selection was greater for heritage speakers who: (1) used
the heritage language more on average, (2) used the heritage language in more diverse contexts, and (3) felt more identified with the heritage language.

Research question 4: *Are there asymmetries in heritage speakers’ sensitivity to the conditioning of mood selection in comprehension vs. production?*

The findings of this study did not provide evidence in support of an asymmetry in heritage speakers’ sensitivity to the conditioning of mood selection in comprehension versus production. In fact, participants displayed robust, systematic and nuanced knowledge of the morphosyntactic intricacies of mood in Spanish in both modalities; that is, all constraints (lexical conditioning, structural conditioning, morphological regularity and lexical frequency) were significant both in the auditory pupillometry and elicited production experiments.

Research question 5: *Does heritage speakers’ sensitivity to the lexical and structural factors that constrain mood diverge from that of Spanish-dominant bilinguals of the target language?*

The results of the present study suggest convergence between the heritage and Spanish-dominant samples. Both groups showed nuanced sensitivity to the lexical and structural factors that constrain mood selection in Spanish. In addition, similar effects of regularity and frequency were found both in comprehension and production for the heritage and Spanish-dominant samples. For both groups, the subjunctive was easier to process with morphologically irregular subordinate verbs and higher frequency governors. In addition, the magnitude of the regularity and frequency effects was greater in lexically conditioned contexts for both groups.
In all, the results reviewed so suggest that heritage speakers can display native-like knowledge of the Spanish subjunctive. A novelty of the present study, and one that advances the field of heritage language studies, lies in that a comprehensive approach was taken to examine the linguistic knowledge of heritage speakers, ensuring a tight control over the demographics of participants, experimental design and the methodology used. Not only does this comprehensive and controlled approach lend support to the robust validity of the results reported in the present dissertation, it also paves the way to a new way of studying heritage grammars. In addition, it fundamentally challenges prescriptivist notions that are pervasive in the field of heritage language research. In the following subsection, a critical comment on the shortcoming of prescriptivist concepts is provided.

Dissecting incomplete acquisition: problematic definitions and quantifications

When compared to monolingual or Spanish-dominant native speakers, heritage speakers—and also L2 learners—are usually described in terms of being non-native-like, their grammars often being deemed deficient. However, throughout this dissertation, I have discussed literature showing that even when they diverge from established baselines, heritage speakers do have robust and systematic knowledge of the mood system in Spanish. The fact that their knowledge is variable does not imply that is incomplete or deficient; rather, this is a reflection of their highly variable bilingual experience and their community of membership. Accepting that variability is a natural feature of language knowledge and that variable performance needs not be equated with errors leads to reconsider concepts such as incomplete acquisition, attrition, fossilization and ultimate attainment.
For one, the concept of incomplete acquisition is grounded on the notion that there is a kind of idealized, ultimate attainment stage, homogenously reached by native speakers, that somehow represents a complete linguistic target to be acquired. However, characterizing a linguistic system as complete leads two major dilemmas: how to define and how to quantify completeness. Regarding definition, it seems that complete equates a certain prescriptive standard language spoken by an idealized native speaker population that shares an identical mental grammar:

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. (Chomsky 1965: 3)

However, there is ample variation in native speakers’ L1 knowledge. For instance, Dąbrowska and Street (2006) showed that native speakers of English differed in their comprehension of different types of active and passive sentences (plausible vs. implausible) according to their education level, such that those with lower education levels were worse at comprehending passives, especially implausible ones. They concluded that less educated speakers had had overall less experience with passive constructions, which are even less frequent in social situations where colloquial speech is used. Thus, participants experienced comprehension difficulties because reduced use of this structure led to it not being strongly entrenched in their linguistic knowledge (see also Dąbrowska, 2008, 2012; Street & Dąbrowska, 2010)
Concerning quantification, the notion of completeness entails that there must be a quantifiable amount that is necessary to achieve complete, “native-like performance”; that is, what would be sufficient input to safely assume that a child has successfully acquired her L1? This theory fails to provide information on how to measure this amount of input, as well as to define what amount would be crucial. Input quantity is difficult to measure, especially in heritage settings, which are naturalistic as opposed to formal. While empirical research suggests that even when relative frequency of exposure amounts to no more than 30%, this need not lead to divergent attainment (Meisel, 2019b), concluding that bilinguals receive less input because there are only so many hours in a day is simplistic Meisel (2019a). The number of hours spent with speakers of a language does not necessarily represent reliable information about quantity of exposure, as studies have shown that adults vary enormously in the amount of speech they produce when interacting with children (de Houwer, 2014; Weisleder & Fernald; 2013). What is more, it must be considered that not all input becomes intake, that is, in order for children to actually learn from their input, this must be child-directed speech rather than mere exposure to the language (Carroll, 2007). Another problem with the idea behind the notion of input reduction is grounded on the L1 acquisition literature: there is ample evidence that even under less-than-ideal input conditions, children are expert recognizers (Hudson-Kam & Newport, 2005; Meisel, 2007); in fact, they are more effective than adults at constructing grammatical paradigms from linguistic input regardless of the level and/or quality of input (Putnam & Sánchez, 2013: 486).

In addition to these fundamental flaws, calling heritage grammars incomplete undermines the fact that these grammars are organized and systematic (Polinsky, 2018). As stated by Kupisch and Rothman (2016) “grammars that are sufficiently developed for communication cannot be incomplete […] difference from an arbitrary standard does not
suggest that said system is incomplete” (p.10). Using the term incomplete has inevitable consequences beyond science. It has been argued that the term is not only theoretically flawed, but also potentially harmful given its negative connotations. In light of this, some researchers are advocating for the use of the term *divergent attainment* (see Otheguy, 2013, and Polinsky, 2018 for an in depth-discussion), which somewhat more neutrally refers to the differences between the grammars of heritage speakers and other speaker groups.

The findings of this dissertation are in line with those of Bookhammer (2013) and LaCasse (2018), among others, showing that participants in the current heritage sample displayed systematic knowledge of the morphosyntactic intricacies of mood in Spanish both in production and comprehension. Heritage speakers’ performance showed the same trends as the Spanish-dominant bilingual control group such that both samples were sensitive to the same constraints (lexical and structural), and the same lexical factors modulated their knowledge of mood selection (morphological regularity and lexical frequency). In addition, heritage speakers knowledge of the conditioning of subjunctive selection was highly nuanced: in line with the control group, the heritage group showed greater sensitivity to the *lexical* conditioning of mood selection overall (see Chapter 3 subsection *Sensitivity to the Lexical vs. Structural Conditioning of mood selection* for an in depth discussion of why this might be the case) and lexical factors influenced lexically selected subjunctive more strongly in their production and comprehension as well.

These results provide robust evidence that when (1) the heritage population is appropriately characterized and (2) the methodology and materials used are appropriate as well, heritage speakers’ knowledge of the subjunctive may not differ from Spanish-dominant bilinguals, even in the face of the intrinsic difficulties in the acquisition of mood selection and the variability inherent to heritage grammars.
Usage-based approaches to language: moving beyond prescriptive notions in the study of heritage populations

Despite the fact that heritage grammars have been extensively researched, consensus is still far from being reached on how to appropriately characterize them and their speakers. Given that applying prescriptive notions to the study of heritage grammars has not yielded optimal results, it is important for usage to start taking a central role in theories of heritage language acquisition and processing. Because usage-based approaches place a strong emphasis on the effects of life-long experience with language use, they are an exceptionally well-suited tool to understand the process of language acquisition. Thus, these approaches to language can contribute to the field of heritage language research by providing a more robust account of how different amounts of exposure and use, as well as their interaction with the nature and diversity of the social context shape heritage grammars.

Usage-based approaches propose that linguistic knowledge is shaped by the everyday interaction of language use with domain-general cognitive mechanisms. In addition, these approaches emphasize that linguistic knowledge is not just shaped by linguistic but also social factors, as it is “grounded in and emergent from language use in concrete social activity for specific purposes that are tied to specific communities of practice” (Hall et al. 2006: 235). Therefore, usage-based approaches to language offer a nuanced account of the relationship between experience and acquisition, detailing how experience dynamically shapes usage and processing. This view shifts the focus away from whether there is sufficient experience for a given structure to be acquired—or to prevent its attrition—asking instead what structure we can expect to emerge given different amounts and kinds of experience, and how changes in the frequency of specific structures can lead to a gradual evolution of grammatical patterns over time. Thus, these
approaches offer a unbiased and principled way of understanding the origins of what might, on the surface, appear to be incomplete acquisition or attrition by examining the processes that shape heritage grammars (Putnam and Sanchez 2013; Rothman 2009; Flores 2015; Yang 2016).

In addition to this, the emphasis that usage-based approaches place on understanding and accounting for variability within and across individuals makes them ideal for the study of such a highly heterogeneous population as heritage speakers. Under a usage-based approach, the fact that heritage speakers’ knowledge can be more variable than that of other populations does not entail in any way that is incomplete or deficient, rather, this reflects their highly variable bilingual experience and their community of membership. Accepting that variability is a natural feature of language knowledge and that variable performance need not be equated with errors, leads to reconsider concepts such as incomplete acquisition, attrition, fossilization and ultimate attainment. Importantly, this conceptual innovation calls for several methodological changes in how study heritage populations are studied. Here, I will discuss four core aspects: (1) using online tasks with ecologically valid materials, (2) appropriately selecting a heritage speaker group and a benchmark group, (3) using production and comprehension tasks complementarily, and (4) exploring both linguistic and sociolinguistic factors.

First, an ample body of research has shown that heritage speakers have difficulty with explicit tasks that require the use of meta-linguistic knowledge (Montrul & Bowles, 2009), and that oftentimes the input this population receives is quantitatively and qualitatively different from that of monolingual speakers (Pires & Rothman, 2009). For instance, because grammaticality judgements assume a scale that is measured in terms of a monolingual L1 grammar, heritage speakers’ judgements will necessarily differ from those of monolingual speakers by virtue of their linguistic experience being different,
which makes applying the same scales to both populations invalid (Geeslin, 2003; Geeslin & Gudmestad, 2008; Geeslin & Guijarro-Fuentes, 2006; Gudmestad, 2006). In relation to this, Otheguy (2013) argues that “when experimental subjects are tested on the basis of items belonging to somebody else’s grammar, their lack of success (which leads to the conclusion of incompleteness) is preordained and unavoidable, and therefore unrevealing” (p. 4). Given this, it is necessary to create materials that are ecologically valid and that reflect the socio-dialectal features of heritage speakers’ communities of practice rather than standardized norms selected by researchers. Furthermore, these materials should then be implemented in experiments using techniques such as eye-tracking, pupillometry or ERPs that collect implicit measures of knowledge and performance (see Murahata, Murahata & Cook, 2016). Crucially, using real-time, online processing measures can help reveal convergence that is undetectable in offline tasks, as shown in the present dissertation (Boxell & Felser, 2017; Felser & Cunnings, 2012).

Second, the heritage speakers should be chosen with careful regard of their community of practice, given that this construct will directly impact the predictions and hypothesis that can be made on the sample’s performance and knowledge. As shown in this dissertation, the community that the heritage speakers belongs to will affect the results observed. In this case, New Mexico is the longest established bilingual community in the United States, so it is not surprising that the minority language has remained vital and, therefore, heritage speakers largely converged with the Spanish-dominant bilinguals in the control group. This leads to the next point, which is the selection of an appropriate comparison group. Traditional, studies have compared heritage speakers to monolingual speakers or native speakers. While heritage speakers are indeed native speakers of the heritage language (see Rothman & Treffes-Daller, 2014, for a lengthy discussion), they have radically different linguistic experiences from monolinguals.
(environment, socialization, bilingualism and literacy among others), which makes this comparison largely uninformative.

Third, using production and comprehension tasks complementarily improves our understanding of heritage speakers’ underlying grammatical representations. There is in fact an emergent trend within linguistics that calls for the need to triangulate experimental data using a wide range of elicitation techniques. There is strong evidence in favor of production-comprehension links in a wide variety of linguistic phenomena (Beatty-Martínez & Dussias, 2017; Christiansen & Chater, 2016; Guzzardo Tamargo, Valdés Kroff & Dussias, 2016; Lozano & Mendikoetxea, 2018; McDonald, 2013; Pickering & Garrod, 2007, 2013; inter alia). The cognitive mechanisms that guide the consolidation of linguistic knowledge modulate the linguistic choices that speakers make during production which, over time, generate distributional patterns in the language input that comprehenders are exposed to. In turn, these distributional patterns become the probabilistic constraints that guide comprehension (Bybee 2006, 2013; Goldberg 2006; Hopper 1979, 1988; MacDonald 1999, 2013). For instance, Beatty-Martínez and Dussias (2017) tested two groups of Spanish–English bilinguals with different codeswitching experiences; one group consisted of bilinguals who were habitual code-switchers in their community, whereas the other rarely code-switched. The results of their reading experiment utilizing event related potentials (ERPs henceforth) showed that bilinguals who were habitual code-switchers were sensitive to violations in the distributional patterns of their everyday switching preferences, so that an incongruent switch such as *la fork (where fork is masculine in Spanish and should therefore be preceded by the masculine determiner el) produced an N400 modulation indicating semantic integration difficulties. By contrast, bilinguals who were not habitual code-switchers were insensitive to this type of violation. By better understanding both the production and comprehension
skills of heritage speakers, scholars can achieve a more holistic understanding of speakers’ knowledge of the distributional regularities of the heritage language. In addition, while some populations of heritage speakers may not present asymmetries in their production and comprehension skills, as is the case in the present study, other groups have shown decreased production abilities. Studying this asymmetry can shed light on how knowledge of the heritage grammar is acquired over time, placing emphasis on the process rather than just the result (Putnam & Sánchez, 2013).

Fourth and final, going beyond linguistic factors is crucial given that it has been shown that linguistic variation is modulated by social factors (Labov, 1972, 1984). When a study only investigates linguistic factors, the theory resulting from it is necessarily incomplete and partial. Because languages are social constructs, linguistic knowledge cannot be eviscerated from its sociolinguistic context. Therefore, any explanatorily adequate theory of heritage languages needs to take sociolinguistic factors into full consideration. This is especially relevant given that there is strong evidence that psychosocial variables and biographical factors robustly impact heritage grammar outcomes (see: Ducar, 2012, for motivation; Allard & Landry, 1992, for language attitudes; Rivera-Mills, 2012, for generation; Swender, Martin, Rivera-Martínez & Kagan, 2014, for years of formal education in the heritage language). As of now, more evidence is necessary to elucidate which sociolinguistic variables affect different aspects of heritage grammar acquisition and variability and how.

Quo Vadis? Future Directions
As a relatively young area of research, the field of heritage studies is rife with possibilities for new and innovative studies. In following up with the current study my goal is
twofold: on the one hand it would be interesting to study a new set of governors and, on the other hand, a new set of speakers.

Thus, a first goal would be to study heritage speakers’ processing of variable governors. As a reminder, variable governors are those lexical items that do not select the subjunctive categorically; that is, their rate of subjunctive selection can vary, being as low as 3% or as high as 80%. Variable governors make it possible to investigate the interaction between rate of subjunctive selection of the governor (a lexical constraint) with negation (a structural constraint). Because prior corpus studies have shown that structural constraints (operating on subjunctive selection are not as strong as the lexical constrains (Torres Cacoullos et al., 2017), it is possible that a variable governor such as pensar (‘to think’)–which selects the subjunctive mood 33% of the time–might take precedence over negation, which strongly favors subjunctive selection. Thus, a subordinate verb in the subjunctive mood preceded by a negated governor, may or may not be processed with greater ease than one in the indicative.

A goal is to study variation in a group of monolingual Spanish native speakers. Pupillometry has yet to be applied to the study the subjunctive in such a group and, therefore, it is critical to count with such data as a starting point for future theories of mood selection in bilingual populations. While comparison of bilingual populations with monolinguals is far from ideal, it is necessary to establish a prior scientific knowledge of what monolinguals do to better understand more complex speaker groups such as bilinguals in which a higher number of variables complicate their study. In addition, it would be indeed extremely interesting to study a group of monolinguals which was not from a Western, educated, industrialized, rich, and democratic (WEIRD) sample. These individuals represent as much as 80% of study participants but only 12% of the world’s population and might be skewing our understanding of human behavior and culture.
(Henrich, Heine & Norenzayan, 2010; Polinsky, 2018: 27). Studies have shown that variation that is typical of colloquial language registers is preserved and even amplified in heritage speakers’ speech because they are predominantly exposed to colloquial registers and have reduced access to formal instruction in the heritage language and to standard registers […] In this case, heritage speakers may differ only from educated monolinguals, but not from non-educated speakers who, like themselves, are mainly exposed to non-standard registers (Flores, 2019: 30; Rinke & Flores, 2014; Rinke, Flores & Barbosa, 2018). Testing of non-WEIRD populations is paramount to test this extremely interesting hypothesis.

**Conclusion**

The study of heritage populations gives us unique insight into not only the linguistic and cognitive reality of bilingualism, but also its social facets. It is scientifically and socially relevant to disseminate the findings of basic science to contribute to the validation of heritage speakers’ language practices, which are commonly looked down on and stigmatized both by the public and the academic community by virtue of being a minority. While this dissertation is but a small grain of sand, its goal is to view heritage grammars holistically, thereby contributing to a more unbiased theory of language acquisition and processing in heritage language contexts that has greater explanatory adequacy.
APPENDIX: Model outputs

Output for models investigating the effect of morphological regularity and lexical frequency in Lexical and Structural Conditioning.

**Table A.** Model output – Regularity effect on Lexical Conditioning

*Non-Variable Subjunctive Irregular Reference Level*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVS Irreg.)</td>
<td>-22.91</td>
<td>6.26</td>
<td>-3.66</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Smooth terms

<table>
<thead>
<tr>
<th></th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVS Irreg.)</td>
<td>8.50</td>
<td>10.65</td>
<td>4.22</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Diff. NVI Irreg. minus NVS Irreg.</td>
<td>2.01</td>
<td>2.02</td>
<td>1.99</td>
<td>.08</td>
</tr>
<tr>
<td>Diff. NVS Reg. minus NVS Irreg.</td>
<td>2.00</td>
<td>2.00</td>
<td>1.55</td>
<td>.27</td>
</tr>
<tr>
<td>Interaction NVI/NVS by Reg./Irreg.</td>
<td>11.01</td>
<td>13.81</td>
<td>4.20</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.6</td>
<td>28.81</td>
<td>296.99</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>184.35</td>
<td>330.00</td>
<td>1.58</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>240.37</td>
<td>610.00</td>
<td>1.14</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .14; Deviance Explained: 14.70%

**Table B.** Model output – Regularity effect in Lexical Conditioning

*Non-Variable Indicative Regular Reference Level*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI Reg.)</td>
<td>-20.81</td>
<td>4.59</td>
<td>-4.53</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Smooth terms

<table>
<thead>
<tr>
<th></th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI Reg.)</td>
<td>8.539</td>
<td>10.64</td>
<td>5.86</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Diff. NVS Reg. minus NVI Reg.</td>
<td>12.16</td>
<td>14.96</td>
<td>4.16</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Diff. NVI Irreg. minus NVI Reg.</td>
<td>2.411</td>
<td>2.58</td>
<td>1.56</td>
<td>.10</td>
</tr>
<tr>
<td>Interaction NVS / NVI by Irreg./Reg.</td>
<td>2.66</td>
<td>3.05</td>
<td>9.32</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.6</td>
<td>28.81</td>
<td>299.05</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>183.20</td>
<td>329.00</td>
<td>1.52</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>242.82</td>
<td>609.00</td>
<td>1.165</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .10; Deviance Explained: 9.81%
### Table C. Model output – Regularity effect on Structural Conditioning
**Negated Subjunctive Irregular Reference Level**

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-S Irreg.)</td>
<td>-31.24</td>
<td>4.63</td>
<td>-6.75</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-S Irreg.)</td>
<td>7.84</td>
<td>9.91</td>
<td>5.29</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Diff. Neg-I Irreg. minus Neg-S Irreg.</td>
<td>2.76</td>
<td>3.24</td>
<td>3.13</td>
<td>.02</td>
</tr>
<tr>
<td>Diff. Neg-S Reg. minus Neg-S Irreg.</td>
<td>2.00</td>
<td>2.00</td>
<td>0.13</td>
<td>.87</td>
</tr>
<tr>
<td>Interaction Neg-I/Neg-S by Reg./Irreg.</td>
<td>2.00</td>
<td>2.01</td>
<td>1.30</td>
<td>.27</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.93</td>
<td>28.89</td>
<td>396.05</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>202.31</td>
<td>330.00</td>
<td>2.00</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>193.48</td>
<td>578.00</td>
<td>0.85</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .08; Deviance Explained: 8.48%

### Table D. Model output – Regularity effect on Structural Conditioning
**Negated Indicative Regular Reference Level**

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-I Reg.)</td>
<td>-23.36</td>
<td>4.67</td>
<td>-5.00</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI Reg.)</td>
<td>8.04</td>
<td>10.16</td>
<td>5.69</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Diff. Neg-I Reg. minus Neg-S Reg.</td>
<td>2.00</td>
<td>2.00</td>
<td>0.76</td>
<td>.46</td>
</tr>
<tr>
<td>Diff. Neg-I Irreg. minus Neg-I Reg.</td>
<td>2.05</td>
<td>2.07</td>
<td>0.69</td>
<td>.48</td>
</tr>
<tr>
<td>Interaction Neg-S/Neg-I by Irreg./Reg.</td>
<td>3.00</td>
<td>3.50</td>
<td>0.95</td>
<td>.39</td>
</tr>
<tr>
<td>Gaze Y by Gaze Y</td>
<td>27.91</td>
<td>28.88</td>
<td>396.87</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>202.30</td>
<td>330.00</td>
<td>2.00</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>194.11</td>
<td>579.00</td>
<td>0.85</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .08; Deviance Explained: 8.49%
Table E. Model output – Frequency effect on Lexical Conditioning

*Non-Variable Governors*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI)</td>
<td>-28.64</td>
<td>4.98</td>
<td>-5.76</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>NVS</td>
<td>-8.21</td>
<td>3.89</td>
<td>-2.11</td>
<td>0.03</td>
</tr>
</tbody>
</table>

**Smooth terms**

<table>
<thead>
<tr>
<th>Interaction Time by Frequency in NVI</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.74</td>
<td>10.17</td>
<td>5.75</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interaction Time by Frequency in NVS</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze X:Gaze Y</td>
<td>27.58</td>
<td>28.81</td>
<td>306.47</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>183.87</td>
<td>330.00</td>
<td>1.56</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>256.16</td>
<td>609.00</td>
<td>1.31</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .10; Deviance Explained: 9.78%

Table F. Model output – Frequency effect on Structural Conditioning

*Negated Governors*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-I)</td>
<td>-27.70</td>
<td>4.40</td>
<td>-6.29</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Neg-S</td>
<td>-5.71</td>
<td>3.46</td>
<td>-1.65</td>
<td>.03</td>
</tr>
</tbody>
</table>

**Smooth terms**

<table>
<thead>
<tr>
<th>Interaction Time by Frequency in NVI</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.23</td>
<td>7.75</td>
<td>5.66</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interaction Time by Frequency in NVS</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze X by Gaze Y</td>
<td>28.08</td>
<td>28.92</td>
<td>386.12</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>205.07</td>
<td>329.00</td>
<td>2.07</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>194.80</td>
<td>578.00</td>
<td>0.90</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .08; Deviance Explained: 8.54 %
**Table G.** Model output – Proficiency effect on Lexical Conditioning  
*Non-Variable Governors*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI)</td>
<td>-19.89</td>
<td>4.41</td>
<td>-4.50</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>NVS</td>
<td>-8.03</td>
<td>2.21</td>
<td>-3.62</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>Edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Time by Proficiency in NVI</td>
<td>7.22</td>
<td>7.83</td>
<td>6.00</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction Time by Proficiency in NVS</td>
<td>7.37</td>
<td>7.94</td>
<td>3.98</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.51</td>
<td>28.79</td>
<td>294.63</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>188.72</td>
<td>328.00</td>
<td>1.83</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>247.47</td>
<td>609.00</td>
<td>1.22</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .09; Deviance Explained: 9.61%

**Table H.** Model output – Proficiency effect on Structural Conditioning  
*Negated Governors*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-I)</td>
<td>-26.38</td>
<td>4.04</td>
<td>-6.52</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Neg-S</td>
<td>-5.30</td>
<td>2.33</td>
<td>-2.28</td>
<td>.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>Edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Time by Proficiency in Neg-I</td>
<td>12.45</td>
<td>15.26</td>
<td>3.87</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction Time by Proficiency in Neg-S</td>
<td>7.14</td>
<td>7.74</td>
<td>6.03</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.97</td>
<td>28.90</td>
<td>384.48</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>200.29</td>
<td>328.00</td>
<td>1.93</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>193.57</td>
<td>579.00</td>
<td>0.98</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .08; Deviance Explained: 8.75%
Table I. Model output – Heritage Language Use effect on Lexical Conditioning
Non-Variable Governors

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI)</td>
<td>-14.69</td>
<td>4.19</td>
<td>-3.50</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>NVS</td>
<td>-9.51</td>
<td>2.28</td>
<td>-4.17</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>Edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Time by Proficiency in NVI</td>
<td>14.68</td>
<td>17.24</td>
<td>4.56</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction Time by Proficiency in NVS</td>
<td>6.75</td>
<td>7.37</td>
<td>3.40</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.27</td>
<td>28.73</td>
<td>203.85</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>153.59</td>
<td>268.00</td>
<td>1.70</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>200.78</td>
<td>579.00</td>
<td>0.90</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .09; Deviance Explained: 10.00%

Table J. Model output – Heritage Language Use effect on Structural Conditioning
Negated Governors

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-I)</td>
<td>-29.23</td>
<td>4.35</td>
<td>-6.72</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Neg-S</td>
<td>-4.91</td>
<td>2.47</td>
<td>-1.99</td>
<td>.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>Edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Time by Proficiency in Neg-I</td>
<td>9.72</td>
<td>11.84</td>
<td>3.65</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction Time by Proficiency in Neg-S</td>
<td>6.83</td>
<td>7.45</td>
<td>5.00</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.36</td>
<td>28.76</td>
<td>268.23</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>162.06</td>
<td>269.00</td>
<td>1.80</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>179.53</td>
<td>550.00</td>
<td>0.95</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .08; Deviance Explained: 8.68%
Table K. Model output – Language Entropy effect on Lexical Conditioning
Non-Variable Governors

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI)</td>
<td>-16.18</td>
<td>4.32</td>
<td>-3.74</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>NVS</td>
<td>-8.12</td>
<td>2.28</td>
<td>-3.56</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>Edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Time by Proficiency in NVI</td>
<td>9.45</td>
<td>11.58</td>
<td>4.27</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction Time by Proficiency in NVS</td>
<td>6.83</td>
<td>7.46</td>
<td>3.50</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.24</td>
<td>28.72</td>
<td>199.60</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>157.64</td>
<td>269.00</td>
<td>1.98</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>197.44</td>
<td>580.00</td>
<td>0.98</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .09; Deviance Explained: 9.98%

Table L. Model output – Language Entropy effect on Structural Conditioning
Negated Governors

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-I)</td>
<td>-28.16</td>
<td>4.38</td>
<td>-6.44</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Neg-S</td>
<td>-5.82</td>
<td>2.47</td>
<td>-2.35</td>
<td>.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>Edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Time by Proficiency in Neg-I</td>
<td>13.05</td>
<td>15.53</td>
<td>3.47</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction Time by Proficiency in Neg-S</td>
<td>7.10</td>
<td>7.67</td>
<td>6.17</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.34</td>
<td>28.76</td>
<td>270.58</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>161.34</td>
<td>268.00</td>
<td>1.89</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>182.02</td>
<td>549.00</td>
<td>0.10</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .08; Deviance Explained: 9.00%
### Table M. Model output – Language Entropy effect on Lexical Conditioning
#### Non-Variable Governors

**Parametric coefficients**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI)</td>
<td>-16.18</td>
<td>4.32</td>
<td>-3.74</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>NVS</td>
<td>-8.12</td>
<td>2.28</td>
<td>-3.56</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

**Smooth terms**

<table>
<thead>
<tr>
<th></th>
<th>Edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Time by Proficiency in NVI</td>
<td>9.45</td>
<td>11.58</td>
<td>4.27</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction Time by Proficiency in NVS</td>
<td>6.83</td>
<td>7.46</td>
<td>3.50</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.24</td>
<td>28.72</td>
<td>199.60</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>157.64</td>
<td>269.00</td>
<td>1.98</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>197.44</td>
<td>580.00</td>
<td>0.98</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .09; Deviance Explained: 9.98%

### Table N. Model output – Language Entropy effect on Lexical Conditioning
#### Negated Governors

**Parametric coefficients**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-I)</td>
<td>-16.18</td>
<td>4.32</td>
<td>-3.74</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Neg-S</td>
<td>-8.12</td>
<td>2.28</td>
<td>-3.56</td>
<td>&lt; .01</td>
</tr>
</tbody>
</table>

**Smooth terms**

<table>
<thead>
<tr>
<th></th>
<th>Edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Time by Proficiency in Neg-I</td>
<td>9.45</td>
<td>11.58</td>
<td>4.27</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction Time by Proficiency in Neg-S</td>
<td>6.83</td>
<td>7.46</td>
<td>3.50</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>27.24</td>
<td>28.72</td>
<td>199.60</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>157.64</td>
<td>269.00</td>
<td>1.98</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>197.44</td>
<td>580.00</td>
<td>0.98</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .09; Deviance Explained: 9.98%
**Table O.** Model output – Regularity effect on Lexical Conditioning

*Non-Variable Subjunctive Irregular Reference Level*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVS Irreg.)</td>
<td>-36.934</td>
<td>5.183</td>
<td>-7.126</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVS Irreg.)</td>
<td>8.944</td>
<td>11.155</td>
<td>2.321</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Diff. NVI Irreg. minus NVS Irreg.</td>
<td>2.002</td>
<td>2.003</td>
<td>2.205</td>
<td>.11</td>
</tr>
<tr>
<td>Diff. NVS Reg. minus NVS Irreg.</td>
<td>2.001</td>
<td>2.002</td>
<td>1.617</td>
<td>.19</td>
</tr>
<tr>
<td><strong>Interaction NVI/NVS by Reg./Irreg.</strong></td>
<td><strong>9.263</strong></td>
<td><strong>11.649</strong></td>
<td><strong>3.327</strong></td>
<td>&lt; <strong>.001</strong></td>
</tr>
</tbody>
</table>

Diff. NVI Reg. minus NVI Reg. | 25.917 | 28.209 | 319.503 | < .001 |

Diff. NVI Irreg. minus NVI Reg. | 97.423 | 150    | 4.526  | < .001 |

Time by Participant | 217.53  | 619    | 1.057  | < .001 |

Adjusted R-squared: .10; Deviance Explained: 11.30%

**Table P.** Model output – Regularity effect in Lexical Conditioning

*Non-Variable Indicative Regular Reference Level*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI Reg.)</td>
<td>-33.99</td>
<td>5.618</td>
<td>-6.05</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI Reg.)</td>
<td>11.38</td>
<td>13.76</td>
<td>2.40</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Diff. NVS Reg. minus NVI Reg.</strong></td>
<td><strong>5.84</strong></td>
<td><strong>7.42</strong></td>
<td><strong>3.23</strong></td>
<td><strong>.001</strong></td>
</tr>
<tr>
<td>Diff. NVI Irreg. minus NVI Reg.</td>
<td>2.00</td>
<td>2.00</td>
<td>0.49</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Interaction NVS/NVI by Irreg./Reg.</strong></td>
<td><strong>5.35</strong></td>
<td><strong>6.68</strong></td>
<td><strong>4.89</strong></td>
<td>&lt; <strong>.001</strong></td>
</tr>
</tbody>
</table>

Gaze X by Gaze Y | 27.04  | 28.64  | 524.05 | < .001 |

Time by Participant | 112.91 | 149.00 | 4.86  | < .001 |

Time by Item | 393.10  | 618.00  | 2.45  | < .001 |

Adjusted R-squared: .08; Deviance Explained: 9.38%
### Table Q. Model output – Regularity effect on Structural Conditioning

*Negated Subjunctive Irregular Reference Level*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-S Irreg.)</td>
<td>-34.423</td>
<td>3.866</td>
<td>-8.90</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-S Irreg.)</td>
<td>5.01</td>
<td>6.07</td>
<td>2.41</td>
<td>.02</td>
</tr>
<tr>
<td>Diff. Neg-I Irreg. minus Neg-S Irreg.</td>
<td>2.00</td>
<td>2.00</td>
<td>2.86</td>
<td>.05</td>
</tr>
<tr>
<td>Diff. Neg-S Reg. minus Neg-S Irreg.</td>
<td>3.88</td>
<td>4.43</td>
<td>1.79</td>
<td>.12</td>
</tr>
<tr>
<td>Interaction Neg-I/Neg-S by Reg./Irreg.</td>
<td>2.00</td>
<td>2.01</td>
<td>37.30</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>24.96</td>
<td>27.77</td>
<td>143.18</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>72.32</td>
<td>149.00</td>
<td>3.36</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>263.64</td>
<td>639.00</td>
<td>2.76</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .15; Deviance Explained: 12.9%

---

### Table R. Model output – Regularity effect on Structural Conditioning

*Negated Indicative Regular Reference Level*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-I Reg.)</td>
<td>-37.495</td>
<td>4.136</td>
<td>-9.07</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smooth terms</th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI Reg.)</td>
<td>5.79</td>
<td>7.09</td>
<td>3.17</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Diff. Neg-I Reg. minus Neg-S Reg.</td>
<td>2.00</td>
<td>2.01</td>
<td>39.82</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Diff. Neg-I Irreg. minus Neg-I Reg.</td>
<td>3.18</td>
<td>3.54</td>
<td>3.04</td>
<td>.06</td>
</tr>
<tr>
<td>Interaction Neg-S/Neg-I by Irreg./Reg.</td>
<td>2.00</td>
<td>2.00</td>
<td>35.19</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Gaze Y by Gaze Y</td>
<td>24.89</td>
<td>27.73</td>
<td>142.65</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>72.42</td>
<td>150.00</td>
<td>3.38</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>265.73</td>
<td>640.00</td>
<td>2.87</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .15; Deviance Explained: 13%
**Table S.** Model output – Frequency effect on Lexical Conditioning  
*Non-Variable Governors*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept NVI)</td>
<td>-32.39</td>
<td>4.36</td>
<td>-7.43</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>NVS</td>
<td>-10.90</td>
<td>2.62</td>
<td>-4.17</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

**Smooth terms**

<table>
<thead>
<tr>
<th></th>
<th>Edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Time by Frequency in NVI</td>
<td>6.28</td>
<td>7.07</td>
<td>4.33</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction Time by Frequency in NVS</td>
<td>4.58</td>
<td>5.10</td>
<td>1.50</td>
<td>.18</td>
</tr>
<tr>
<td>Gaze X:Gaze Y</td>
<td>25.91</td>
<td>28.20</td>
<td>319.56</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>108.14</td>
<td>149.00</td>
<td>5.34</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>223.99</td>
<td>618.00</td>
<td>1.21</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .1; Deviance Explained: 11.6%

**Table T.** Model output – Frequency effect on Structural Conditioning  
*Negated Governors*

<table>
<thead>
<tr>
<th>Parametric coefficients</th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept Neg-I)</td>
<td>-36.48</td>
<td>3.88</td>
<td>-9.40</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Neg-S</td>
<td>7.34</td>
<td>1.77</td>
<td>4.15</td>
<td>.03</td>
</tr>
</tbody>
</table>

**Smooth terms**

<table>
<thead>
<tr>
<th></th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Time by Frequency in NVI</td>
<td>13.53</td>
<td>15.86</td>
<td>4.88</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction Time by Frequency in NVS</td>
<td>5.74</td>
<td>6.37</td>
<td>3.27</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Gaze X by Gaze Y</td>
<td>24.94</td>
<td>27.76</td>
<td>140.85</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Participant</td>
<td>76.24</td>
<td>149.00</td>
<td>3.57</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time by Item</td>
<td>269.69</td>
<td>638.00</td>
<td>3.08</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R-squared: .16; Deviance Explained: 13.20 %


De Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. Behavioral Sciences, 9, 33.


Di Franco, G. (2016). Multiple correspondence analysis: one only or several techniques?. Quality & Quantity, 50, 1299-1315.


Murphy, M. D. (2008). The role of typological drift in the development of the romance subjunctive: a study in word-order change, grammaticalization and synthesis. University of Texas, Austin: University of Texas at Austin doctoral dissertation.


VITA

PRISCILA LÓPEZ-BELTRÁN

EDUCATION


2009 –2013 Universidad de Málaga (Spain) B.A. in English Studies.

PUBLICATIONS


AWARDS

2019 NSF Dissertation Improvement Award

2018 Adele Miccio Memorial Travel Award

2017 PIRE (Partnerships for International Research and Education) Summer Award

2017 Teaching with Technology Certificate

2016-2018 Foreign Fulbright Scholarship recipient