© 2017
David Giancaspro
ALL RIGHTS RESERVED

HERITAGE SPEAKERS' PRODUCTION AND COMPREHENSION OF LEXICALLY- AND CONTEXTUALLY SELECTED SUBJUNCTIVE MOOD MORPHOLOGY

By
DAVID GIANCASPRO

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Spanish

Written under the direction of

Dr. Liliana Sánchez

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2017

ABSTRACT OF THE DISSERTATION

Heritage Speakers' Production and Comprehension of Lexically- and Contextually-Selected Subjunctive Mood Morphology

By DAVID GIANCASPRO

Dissertation Director:

Liliana Sánchez

This dissertation explores heritage speakers' (HSs) production and comprehension of lexically- and contextually-selected subjunctive mood morphology in Spanish. Unlike Spanish-dominant speakers, whose knowledge of subjunctive mood appears to be largely categorical, HSs' subjunctive mood knowledge is characterized by great variability, both in production and comprehension. In production, HSs have been shown to alternate, in expected subjunctive contexts, between target subjunctive and non-target indicative mood morphology (e.g., Bookhamer, 2013; Giancaspro, under revision; Martillo Viner, 2017; Montrul, 2009; van Osch & Sleeman, *forthcoming*; Perez-Cortes, 2016, Silva-Corvalán, 1994; *inter alia*). In comprehension, HSs have been shown to exhibit less sensitivity to differences in meaning between subjunctive and indicative mood (e.g., Montrul, 2009; Montrul & Perpiñán, 2011; Perez-Cortes, 2016). Though these and other studies have now documented HSs' variable subjunctive knowledge, the underlying nature of this

morphological variability remains unclear. The present study contributes to our understanding of HSs' subjunctive knowledge, and HS variability more generally, by identifying a number of between- and within-group factors that shape HSs' likelihood of producing and understanding subjunctive mood morphology in Spanish.

In the present dissertation project, 42 HSs (22 advanced proficiency, 20 intermediate proficiency), 19 late-childhood immigrants (LCIs) and 20 Spanish-dominant controls (SDCs) completed three experimental tasks assessing their knowledge of (a) intensional (lexically-selected) subjunctive mood with *para que* and (b) polarity (contextually-selected) subjunctive mood in adjectival relative clauses. Results of the three experimental tasks reveal that many HSs, despite making the same grammatical distinctions as the SDCs, exhibit variability in production, comprehension, and preference of subjunctive mood morphology in Spanish. Analyses of between-group factors demonstrated that HSs with *lower* Spanish proficiency and *earlier* age of acquisition of English (AofA Eng) are most likely to perform variably with subjunctive mood. Analyses of within-group factors revealed that HSs are most likely to demonstrate subjunctive mood variability with polarity mood selection, as well as with less frequent Spanish verbs.

These findings highlight the heritage language lexicon as a primary source of HS variability and challenge the practice of classifying HSs dichotomously, that is to say, as having either *acquired* or *not acquired* linguistic properties such as subjunctive mood. Instead, it is argued that many HSs maintain systematic knowledge of the morphosyntactic features that underlie subjunctive mood morphology yet, over time, become less likely to apply and/or recognize these features with less common, lower frequency lexical items.

ACKNOWLEDGEMENTS

I cannot possibly thank, in the space of a few paragraphs, all of the people whose love, support and encouragement have empowered me to finish this dissertation.

I'm exceedingly grateful, first and foremost, for my academic advisors Jason Rothman (MA) and Liliana Sánchez (PhD), whose boundless energy, enthusiasm and creativity have inspired me to work harder and dig deeper than I ever thought I could.

I'm thankful, also, for the support and assistance of my committee members Jennifer Austin, José Camacho and Alejandro Cuza. Without question, your feedback has improved my writing and helped me to become a better and more thorough scholar.

I'm deeply indebted to the many colleagues and friends, both inside and outside of academia, that I am so extraordinarily lucky to have in my corner. I've always appreciated the strength and wisdom and sense of humor that you share with me, but never more so than this year, which has truly been one of the hardest and most trying of my life.

I'm grateful for my family, for believing in me more than I ever believed in myself. Your confidence and love have helped me become steady, yet flexible, tough and yet understanding. Every day I strive to have the consistently positive impact that you all have had, and continue to have, in your day-to-day interactions with the world around you.

Last, but not certainly not least, I'm thankful for the many heritage speakers who have participated in my research and shown me a glimpse of the countless and beautiful shapes that heritage bilingualism can take. It is my sincere hope that this dissertation, possible only because you shared your Spanish with me, will promote a deeper understanding of what it means to grow up bilingual in the United States.

DEDICATION

To my favorite heritage speakers:
John Giancaspro (Italian) and Carol Kratch (Finnish)

TABLE OF CONTENTS

ABSTRACT…….ii

ACKNOWLEDGEMENTS…….iv

DEDICATION…….v

LIST OF TABLES……...xii

LIST OF FIGURES……..xv

CHAPTER 1: INTRODUCTION…….1

1.1. Opening remarks……1

1.2. Gaps in the research and contributions of this project……...5

1.3. Research questions…….7

1.4. Outline of the remaining chapters……..11

CHAPTER 2: DIVERGENCE AND VARIABILITY IN HSs…….13

2.1. Introduction…….13

2.2. Defining HS differences…….16

2.2.1. HS divergence……..16

2.2.2. HS variability……...22

2.3. Accounts of HS divergence/variability……...27

2.3.1. The Input Quality approach…….28

2.3.2. The Representational Differences approach……34

2.3.3. Activation and Lexicon-Based approaches……..39

2.3.4. Summary of approaches to HS divergence and variability……..48

CHAPTER 3: SUBJUNCTIVE MOOD: SYNTAX AND ACQUISITION…….50

3.1. Introduction…….50

3.2. Types of subjunctive mood…….50

3.2.1. Intensional subjunctive……52

3.2.2. Polarity subjunctive…….58

3.3. Acquisition of subjunctive in Spanish as a native language……...64

3.3.1. Acquisition of subjunctive mood: L1 children……65

3.3.1.1. The L1 learning task…….65

3.3.1.2. Findings from L1 acquisition research…….69

3.3.2. Acquisition of subjunctive mood: 2L1 children……..72

3.3.2.1. The 2L1 learning task……...73

3.3.2.2. Findings from 2L1 acquisition research……...75

3.3.3. Acquisition of subjunctive mood: adult HSs……...79

3.3.3.1. The adult HS learning and maintenance task……79

3.3.3.2. Findings from adult HS research……..82

3.4. Chapter summary……91

CHAPTER 4: METHODOLOGY…….92

4.1. Introduction…….92

4.2. Research questions……..92

4.3. Participants……102

4.3.1. Advanced HSs……103

4.3.2. Intermediate HSs……105

4.3.3. Late childhood immigrants……105

4.3.4. Spanish-dominant controls…….107

4.3.5. Between-group comparisons……..108

4.3.5.1. DELE……..108

4.3.5.2. English self-ratings……109

4.3.5.3. Spanish self-ratings……111

4.3.5.4. Spanish usage and exposure……112

4.3.5.5. Summary of between-group comparisons……..115

4.4. Procedure……..116

4.4.1. DELE proficiency test……117

4.4.2. Contextualized elicited production task (CEPT)……...118

4.4.3. Contextualized acceptability task (CAT)……...128

4.4.4. Mood preference task (MPT)……134

4.5. Chapter summary……..142

CHAPTER 5: RESULTS……144

5.1. Introduction……...144

5.2. Contextualized elicited production task (CEPT)……..144

5.2.1. Descriptive statistics……..145

5.2.1.1. Variables in the CEPT……145

5.2.1.2. Data trimming and exclusions……146

5.2.1.3. Descriptive results……...148

5.2.2. Statistical models……...150

5.2.2.1. Introduction to GLMMs……..151

5.2.3. Results from the GLMMs……..152

5.2.3.1. Group, ExpectedMood and MoodSelectType……152

5.2.3.1.1. Main effects…….153

5.2.3.1.2. Two-way interactions……...155

5.2.3.1.3. Three-way interaction……..158

5.2.3.2. Group, MoodSelectType and Frequency……164

5.2.3.2.1. Main effects…….165

5.2.3.2.2. Two-way interactions……...165

5.2.3.2.3. Three-way interaction……..167

5.2.3.3. Group, MoodSelectType and PrimeStatus……..171

5.2.3.3.1. Main effects…….172

5.2.3.3.2. Two-way interactions……...172

5.2.3.3.3. Three-way interaction……..173

5.2.3.4. Summary of results from the CEPT……175

5.3. Contextualized acceptability task (CAT)……..177

5.3.1. Descriptive statistics……..177

5.3.1.1. Variables in the CAT……..177

5.3.1.2. Descriptive results……...178

5.3.2. Statistical models for the CAT……...180

5.3.2.1. Transforming the raw data……..180

5.3.3. Results from the GLMMs……..181

5.3.3.1. Group, Mood and Scenario…….181

5.3.3.1.1. Main effects…….182

5.3.3.1.2. Two-way interactions……...183

5.3.3.1.3. Three-way interaction……..187

5.3.3.2. Group, Frequency and Mood (Found)……192

5.3.3.2.1. Main effects…….192

5.3.3.2.2. Two-way interactions……..193

5.3.3.2.3. Three-way interaction……..194

5.3.3.3. Group, Frequency and Mood (Wrong)……...198

5.3.3.3.1. Main effects…….198

5.3.3.3.2. Two-way interactions……...198

5.3.3.3.3. Three-way interaction……..199

5.3.4. Summary of results of the CAT…….204

5.4. Mood preference task (MPT)……205

5.4.1. Descriptive statistics……..205

5.4.1.1. Variables in the MPT……..206

5.4.1.2. Descriptive results……..206

5.4.2. Statistical models for the MPT……..208

5.4.2.1. Transforming the raw data……..209

5.4.3. Results from the GLMMs……..209

5.4.3.1. Group, MoodSelectType and ExpectedMood…….210

5.4.3.1.1. Main effects…….210

5.4.3.1.2. Two-way interactions……...211

5.4.3.1.3. Three-way interaction……..215

5.4.4. Summary of results from the MPT……220

5.5. Summary of Chapter 5……..222

CHAPTER 6: DISCUSSION……...223

6.1. HS differences: divergence and variability……...223

6.1.1. HS divergence……224

6.1.2. HS variability…….226

6.2. Between-group factors……..231

6.2.1. Spanish proficiency……232

6.2.2. Age of acquisition of English……239

6.2.3. Summary of between-group factors……..249

6.3. Within-group factors……249

6.3.1. Mood selection type……...250

6.3.2. Lexical frequency……..256

6.3.3. Structural priming……..279

6.3.4. Summary of within-group factors……..284

6.4. Modeling HSs' divergence and variability with subjunctive mood……..284

6.4.1. Tracing subjunctive mood development……284

6.4.2. Four stages of HSs' subjunctive mood knowledge……287

CHAPTER 7: CONCLUSION……291

7.1. Introduction……...291

7.2. Summary of main findings……291

7.3. Limitations and future directions……..299

BIBLIOGRAPHY……303

LIST OF TABLES

Table 1 Mood morphology in Spanish……...51

Table 2 Participants' probability of producing subjunctive……...90

Table 3 Summary of participant groups……..115

Table 4 Polarity mood selection: baseline conditions…….120

Table 5 Polarity mood selection: comparison of baseline and prime conditions……121

Table 6 Intensional mood selection: indicative and subjunctive example items…….122

Table 7 Intensional mood selection: comparison of baseline and prime conditions…..123

Table 8 Summary of experimental conditions in the CEPT……124

Table 9 Summary of experimental conditions in the CAT……..130

Table 10 Summary of filler conditions in the CAT…….132

Table 11 Example item from polarity, +presupposition condition……..136

Table 12 Example item from polarity, -presupposition condition……...137

Table 13 Example item from intensional, *para que* condition……138

Table 14 Example item from intensional, *porque* condition……..139

Table 15 Summary of experimental conditions in the MPT……...139

Table 16 Summary of filler conditions in the MPT……141

Table 17 Summary of participant groups……142

Table 18 Items excluded from CEPT analysis……147

Table 19 Participants' use of subjunctive by Group, Condition……..149

Table 20 Odds of using subjunctive by Group……154

Table 21 Odds of using subjunctive by MoodSelectType……..154

Table 22 Odds of using subjunctive by ExpectedMood…….155

Table 23 Odds of using subjunctive: Group x ExpectedMood……156

Table 24 Odds of using subjunctive: ExpectedMood x MoodSelectType……..158

Table 25 Odds of using subjunctive: Group x MoodSelectType x ExpectedMood……159

Table 26 Odds of using subjunctive: Group x MoodSelectType x Frequency……167

Table 27 Odds of using subjunctive: Group x MoodSelectType x PrimeStatus…….174

Table 28 Average acceptability ratings by Group, Scenario and Mood……..179

Table 29 Average acceptability ratings by Group, Scenario and Aspect……180

Table 30 Probability of acceptance by Group…….182

Table 31 Probability of acceptance by Scenario……..182

Table 32 Probability of acceptance by Mood……..183

Table 33 Probability of acceptance: Group x Scenario……...184

Table 34 Probability of acceptance: Group x Mood……185

Table 35: Probability of acceptance: Scenario x Mood……..186

Table 36 Probability of acceptance: Group x Scenario x Mood…….190

Table 37 Probability of acceptance: Group x Frequency x Mood (Found)……194

Table 38 Probability of acceptance: Group x Frequency x Mood (Wrong)……...200

Table 39 Response Accuracy by Group, MoodSelectType, ExpectedMood…….207

Table 40 Response accuracy by Group, Condition (Fillers)……..208

Table 41 Probability of accuracy by Group……210

Table 42 Probability of accuracy by MoodSelectType……211

Table 43 Probability of accuracy by ExpectedMood……..211

Table 44 Probability of accuracy: Group x ExpectedMood……212

Table 45 Probability of accuracy: Group x MoodSelectType…….214

Table 46 Probability of accuracy: ExpectedMood x MoodSelectType……...215

Table 47 Probability of accuracy: Group x MoodSelectType x ExpectedMood…….216

Table 48 AdvHSs: AofA Eng subgroups…….241

Table 49 IntHSs: AofA Eng subgroups……..241

LIST OF FIGURES

Figure 1 DELE scores by experimental group……109

Figure 2 English self-ratings by experimental group……..110

Figure 3 Spanish self-ratings by experimental group……..111

Figure 4 Participants' reported Spanish use with parents, friends……...113

Figure 5 Participants reported exposure to Spanish *from* parents……114

Figure 6 Example of experimental item from CEPT……..125

Figure 7 Example of experimental item from CAT……131

Figure 8 Example of experimental item from MPT……140

Figure 9 Probability of subjunctive: Group x MoodSelectType x ExpectedMood…….163

Figure 10 Probability of subjunctive: Group x MoodSelectType x Frequency……..170

Figure 11 Probability of subjunctive: Group x MoodSelectType x PrimeStatus……175

Figure 12 Probability of acceptance: Group x Scenario x Mood……191

Figure 13 Probability of acceptance: Group x Frequency x Mood (Found)……197

Figure 14 Probability of acceptance: Group x Frequency x Mood (Wrong)……...203

Figure 15 Probability of accuracy: Group x ExpectedMood x MoodSelectType……..219

Figure 16 HSs' use of subjunctive by proficiency subgroup……...234

Figure 17 AdvHSs' use of subjunctive: AofA Eng x MoodSelectType……..242

Figure 18 IntHSs' use of subjunctive: AofA Eng x MoodSelectType……243

Figure 19 AdvHSs' acceptability ratings by AofA Eng, Condition……244

Figure 20 IntHSs' acceptability ratings by AofA Eng, Condition……..245

Figure 21 HSs' subjunctive use: MoodSelectType x Frequency……257

Figure 22 AdvHSs' probability of acceptance: Condition x Frequency…….258

Figure 23 Four stages of HL activation: Putnam & Sánchez (2013)……...265

Figure 24 AdvHSs divided by stages: intensional subjunctive……267

Figure 25 IntHSs divided by stages: intensional subjunctive……..268

Figure 26 AdvHSs' use of intensional subjunctive: Stage x Frequency……..270

Figure 27 IntHSs' use of intensional subjunctive: Stage x Frequency…….270

Figure 28 AdvHSs' accuracy with intensional subjunctive: Stage x Task……...271

Figure 29 IntHSs' accuracy with intensional subjunctive: Stage x Task…….272

Figure 30 AdvHSs divided by stages: polarity subjunctive…….273

Figure 31 IntHSs divided by stages: polarity subjunctive……..274

Figure 32 AdvHSs' use of polarity subjunctive: Stage x Frequency……..275

Figure 33 IntHSs' use of polarity subjunctive: Stage x Frequency……275

Figure 34 AdvHSs' acceptability ratings: Stage x Condition…….277

Figure 35 IntHSs' acceptability ratings: Stage x Condition……278

Figure 36 HSs' preference accuracy with polarity subjunctive: Stage x Proficiency….279

Figure 37 Conceptualization of HSs' stages of subjunctive mood knowledge……...289

CHAPTER 1: INTRODUCTION

1.1 Opening remarks

As long as humans immigrate from one linguistic region to another, there will be bilinguals presented with the daunting challenge of simultaneously (a) maintaining their home/native language and (b) mastering the societal language of their new home. For those who immigrate in adulthood—henceforth, first-generation immigrants—the task of acquiring a new societal language is often extraordinarily difficult, as has been extensively documented in decades of second language acquisition research (e.g., Bley Vroman, 1990; Johnson & Newport, 1989; Long, 1990; although see Birdsong, 1999, 2009). Not surprisingly then, first-generation immigrants typically maintain a lifelong preference for, as well as greater competence in, their first language (e.g., Taylor, López, Martínez & Velasco, 2012)—even after years in their new linguistic region.

The children of these first-generation immigrants, on the other hand, tend to quickly master the societal language (e.g., Portes & Rumbaut, 2001), largely as result of early and extensive childhood exposure to the societal language in school, media and a plethora of social situations (e.g., Jia & Aaronson, 2003). Though these children, known as second-generation immigrants, will not struggle to acquire the societal language (Portes & Rumbaut, 2001; Rumbaut, Massey & Bean, 2006) in the way that their parents do, they *will* confront the different, and perhaps equally formidable linguistic challenge of first acquiring, and then, maintaining knowledge of a home language that they typically have less opportunities to hear and use on a daily basis.

In the United States, which Rumbaut et al (2006) have gloomily referred to as the "graveyard of languages" (p. 458) because of its ability to "extinguish…mother tongues within a few generations" (p. 458) the language-maintenance task can be even more difficult. When first generation immigrants come to the US from a non-English speaking country, the native language that they bring with them typically disappears, at least at the level of the family, at or before the third generation (e.g., Alba, Logan, Lutz & Stults, 2002; Carreira, 2013; Rumbaut & Massey, 2013; Schecter & Bayley, 2002; *inter alia*). While this strong, and seemingly inevitable shift towards English monolingualism is less pronounced in communities with higher proportions of immigrants (e.g., Alba et al, 2002; Lynch, 1999; Pérez-Leroux, Cuza & Thomas, 2011; Portes & Rumbaut, 2001; Potowski, 2011; Rivera-Mills, 2011), it is still the case that English usually becomes the stronger language for second-generation immigrants in the US, even when they live in dense ethnic enclaves (e.g., Hialeah, FL) and, consequently, have considerable opportunities to hear and use the home language (e.g., Lynch, 1999; Zentella, 1997).

Though second-generation immigrants (and sometimes, their third-generation offspring) have been speaking home languages, alongside societal languages, for hundreds of years—including in the US (Rumbaut & Massey, 2013)—they have only become a population of interest for bilingualism researchers over the last 50 to 60 years, starting with seminal work by Fishman (1964), Cummins (1979), Dorian (1981) and others. During this relatively brief time period, however, these bilinguals have been given many labels in the academic literature, each of which illuminates one of their often-observed characteristics.

Dorian (1981), in her work on Gaelic speakers in English-dominant Scotland, used the term "semi-speaker" to refer to second- and third-generation Gaelic home speakers

whose knowledge of Gaelic is "markedly different from the fluent-speaker norm" (p. 115), e.g., from the patterns produced by first-generation speakers. This term, which has now fallen entirely out of use, was intended to capture the observation that many second- and third-generation Gaelic speakers struggle considerably to *produce* (e.g., speak) Gaelic, though they continue to understand it.

Lipski (1993), in his work on Spanish in the US, introduced the term "transitional bilingual" to refer to home-language speakers of Spanish who undergo a strong shift in relative linguistic dominance, before adolescence, from Spanish to English. This term, though now quite uncommon in academic texts, addresses the observation that many second-generation immigrants start out as very comfortable speakers of their home language during early childhood only to ultimately become more comfortable speakers of the societal language, which receives broader support.

In the last 15 years or so, however, the most accepted term used to refer to bilingual speakers of a home language is "heritage speaker" (henceforth, HS), originally coined by Valdés (2000) in the context of the US and later broadened to include home-language speakers in other societal contexts. In the present dissertation, I follow Montrul (2016a) in defining HSs as "individuals from language minority groups who grow up exposed to a minority language in the home and the majority societal language" (p. 16).

Research on HSs' knowledge of the heritage language (henceforth, HL) has uncovered two prevailing trends that I will call *HS divergence* and *HS variability*. HS divergence is a comparative term referring to those cases in which HSs exhibit knowledge of the HL which is quantitatively and/or qualitatively different from the knowledge demonstrated by dominant-speakers of the HL e.g., first-generation immigrants. HS

variability, on the other hand, refers to those situations in which individual HSs differ from *themselves*, e.g., alternating variably between distinct ("target" and "non-target") forms of a given linguistic property. (See McCarthy, 2008 for a similar definition of variability in L2 acquisition.) In Chapter 2, I will further define and operationalize each term.

What causes HSs to exhibit divergence and variability with properties of the HL? Investigators have proposed a variety of different explanations, three of which I will briefly outline here. The first explanation, which I will call the Input Quality approach (e.g., Pires & Rothman, 2009) attributes some, *though crucially not all*, HS divergence and variability to differences in the quality of input, that is to say, the type of HL input to which HSs are exposed. The second explanation, which I will call the Representational Differences approach (e.g., Montrul, 2002, 2008, *inter alia*), attributes HS divergence and variability to representational differences between the underlying linguistic knowledge of HSs and that of dominant-speakers of the HL. The third explanation, which I will call the Activation/Lexicalist approach (Putnam & Sánchez, 2013), attributes HS divergence and variability to HSs' reduced activation of the HL, which causes HSs to experience difficulty accessing functional features (FFs) in online production of the HL.

The present dissertation project evaluates these three approaches to HS divergence and variability by testing HSs' knowledge of subjunctive mood morphology in Spanish.

Broadly speaking, the exploration of HSs' subjunctive mood knowledge can be broken down into two smaller tasks. The first task is to *document and describe* HSs' subjunctive mood knowledge, using both productive and receptive experimental methodologies. The second task is to *explain* how HSs' knowledge of subjunctive mood is shaped by two types of predictive variables: (i) between-group variables, which seek to

explain differences in performance across experimental groups (e.g., HSs vs. dominant-speakers of the HL) and (ii), within-group variables, which seek to explain differences which are observed within the performance of a single group. In examining both between-group and within-group predictors, I hope to shed light on the factors that shape divergence and variability in HL acquisition and maintenance.

1.2. Gaps in the research and contributions of this project

A large number of previous studies have explored HSs' knowledge of subjunctive mood morphology in Spanish (e.g., Bookhamer, 2013; Lynch, 1999; Martínez-Mira, 2006; Montrul, 2007, 2009; Montrul & Perpiñán, 2011; Ocampo, 1990; van Osch & Sleeman, forthcoming; Pascual y Cabo, Rothman & Lingwall, 2012; Perez-Cortes, 2016). Nonetheless, there are a few important gaps in the research which I seek to address in the present dissertation project.

The first gap in the research is methodological in nature. While almost all studies of HSs' subjunctive mood knowledge include productive tasks, most of these tasks involve spontaneous speech (e.g., Bookhamer, 2013; Lynch, 1999), making it impossible for researchers to systematically manipulate the effects of certain variables on HSs' subjunctive mood production. Only a few studies, most notably Giancaspro (under revision), van Osch & Sleeman (*forthcoming*) and Perez-Cortes (2016), have utilized carefully controlled elicited production tasks to test the effect of different variables on HSs' subjunctive mood knowledge. Of these three studies, only Giancaspro (under revision) and van Osch & Sleeman (*forthcoming*) collected oral production data. The first

methodological contribution of this study, therefore, is its use of a controlled, oral elicited production task to complement the use of judgment/preference data (Ionin & Zyzik, 2014).

Previous research on HSs' knowledge of subjunctive mood morphology has employed a variety of innovative tasks to test HSs' subjunctive comprehension, including sentence conjunction tasks (e.g., Montrul, 2009), truth value judgment tasks (e.g., Perez-Cortes, 2016) and acceptability judgment tasks (e.g., van Osch & Sleeman, *forthcoming*). None of these tasks, however, have included auditory stimuli (e.g., Montrul, Bhatt & Ghirju, 2015) which may be easier for HSs[1]. The second methodological innovation of the present study, therefore, is the use of two different auditory comprehension tasks: The Contextualized Acceptability Task (CAT) and the Mood Preference Task (MPT). By creating comprehension tasks with a naturalistic, auditory component, I hope to give HSs a better chance to exhibit their knowledge of mood morphology in Spanish.

A second gap in the literature is the limited number of explanatory variables which have been explored in research on HSs' knowledge of subjunctive mood. Previous research has identified HL proficiency (e.g., Giancaspro, under revision; Montrul, 2009; Perez-Cortes, 2016), HL usage (Perez-Cortes, 2016) and subjunctive mood type (e.g., Giancaspro, under revision; see Perez-Cortes, 2016 for a different finding) as variables which can significantly impact HSs' subjunctive mood knowledge. While each of these variables has been shown to impact HSs' subjunctive knowledge, it is almost certain that additional, and as yet unexplored, factors also play a considerable role.

---

[1] Up to this point, there is no clear evidence on potential benefits of aural tasks with HSs. Though HSs typically perform better with oral tasks than written tasks (e.g., Bowles, 2011), Montrul, Bhatt & Ghirju (2015) note that HSs' judgments of Spanish differential object marking (DOM) do not seem to be different when the judgment task is written or bimodal (written and aural). It is also possible that adding an aural component to judgment tasks may make HSs more likely to show an acceptance bias since rejecting ungrammatical sentences would mean saying that another native speaker's utterance is not correct.

The present study addresses this possibility by investigating three additional variables which I will briefly summarize here. The first new variable is age of acquisition of English (henceforth, AofA Eng) which has been shown to have a strong impact on HSs' divergence and variability with other linguistic properties, such as preposition placement (Pascual y Cabo & Gómez-Soler, 2015). This variable will allow me to see whether HSs with earlier AofA Eng differ in their subjunctive knowledge from HSs with later AofA Eng. The second new variable is lexical frequency, which will allow me to determine whether HSs are more likely to produce and comprehend subjunctive mood morphology with frequent verbs. Though lexical frequency has not been previously investigated as a variable in HL acquisition research, it has been identified as a strong predictive variable in second language acquistion (SLA) research (see Ellis, 2002 for a review). The third and final new variable, also previously unexplored in HL acquisition research, is structural priming (e.g., Pickering & Ferreira, 2008), which will allow me to test whether HSs are more likely to produce subjunctive mood morphology immediately after previous subjunctive mood exposure. By testing these new variables, I will broaden our understanding of HSs' subjunctive knowledge.

1.3. Research questions

In this section, I will outline the research questions (RQs) of the present study. Because these questions will be presented again in Section 4.2., I will describe them here very briefly, followed by equally short descriptions of the experimental hypotheses.

The first RQ of the present study, presented below in #1, is a descriptive, rather than explanatory question. In order to answer RQ #1, I designed three different

experimental tasks which together assess HSs' knowledge of subjunctive mood. The goal of these tasks, each of which will be illustrated in Chapter 4, is to provide a thorough description of HSs' production, comprehension and preferences with respect to mood morphology in Spanish.

1. What is the nature of HSs' knowledge of lexically (intensional) and contextually (polarity) selected mood morphology in Spanish?

Based on previous research on HSs' knowledge of mood morphology (e.g., Bookhamer, 2013; Martillo-Viner, 2017; Montrul, 2007, 2009; Pascual y Cabo et al, 2012; Perez-Cortes, 2016; *inter alia*) and verbal morphology more generally (e.g., Anderson, 2001; Cuza & Miller, 2015; Dorian, 1981; Gal, 1989; Mikhaylova, 2012; Pascual y Cabo, 2016; Polinsky, 2006; Sherkina-Lieber, 2015; *inter alia*), I hypothesize that HSs will exhibit significant variability in their knowledge of subjunctive mood morphology. The fact that I expect them to evidence variability, however, does not imply that I expect them to lack systematic sensitivity to mood morphology in Spanish. In previous studies (e.g., Montrul, 2009; Perez-Cortes, 2016), HSs exhibiting productive variability of subjunctive mood (e.g., producing indicative in expected subjunctive mood contexts) have nonetheless demonstrated sensitivity to mood distinctions in receptive experimental tasks. It may be the case, therefore, that HSs in the present study will perform more variably with subjunctive mood in productive, rather than receptive, tasks.

The second and third RQs are explanatory in nature, examining different factors which affect HSs' expressed knowledge of subjunctive mood morphology in Spanish.

The second RQ, copied below in #2, addresses the potential impact of between-group factors on HSs' subjunctive mood knowledge. In other words, what factors make some HSs more likely to exhibit variable knowledge of subjunctive mood than other HSs?

> 2. What is the role of between-group factors on HSs' knowledge of subjunctive mood?

The two between-group variables examined in RQ #2 are Spanish proficiency and AofA Eng. Spanish proficiency, as measured by the DELE proficiency test (Bruhn de Garavito, 2002; Duffield & White, 1999; Montrul & Slabakova, 2003), has been found to accurately predict HSs' performance with a number of grammatical properties including subject and object expression (Montrul, 2004b), aspectual morphology (Montrul, 2009; Montrul & Perpiñán, 2011), and mood morphology (Giancaspro, under revision; Montrul, 2009; Perez-Cortes, 2016). It is expected, therefore, that HSs with advanced proficiency in Spanish will exhibit less variability in their knowledge of subjunctive mood than intermediate proficiency HSs.

AofA Eng has also been found to accurately predict HSs' performance with different grammatical properties. Simultaneous HSs, who begin learning the majority language *before* age 3 (e.g., Montrul, 2002), tend to exhibit more HL variability than sequential HSs, defined as HSs who begin learning the majority language *after* age 3. This finding has been reported with aspectual morphology (Montrul, 2002), gender agreement and assignment (Montrul & Potowski, 2007) and prepositional placement (Pascual y Cabo & Gómez-Soler, 2015), among other properties. Though AofA Eng does not always predict HSs' variability in the HL (e.g., Mikulski, 2010; Montrul & Sánchez-Walker, 2013), it is

hypothesized that HSs with earlier acquisition of English will exhibit more subjunctive mood variability in the present study.

The third RQ, copied below in #3, represents a shift in perspective. Instead of comparing differences in the performance of different groups, RQ #3 explores differences in the performance of individual groups. In other words, what linguistic factors make HSs more likely to exhibit variability with certain experimental items/conditions?

3. What is the role of within-group factors on HSs' sensitivity to subjunctive mood?

The three within-group variables examined in RQ #3 are mood selection type, syntactic priming, and lexical frequency. As mentioned above, mood selection type has been a within-group variable in numerous studies of HSs' subjunctive mood, including Giancaspro (under revision), van Osch & Sleeman (*forthcoming*), Pascual y Cabo et al (2012), and Perez-Cortes (2016). It is expected that this variable will have a significant effect in the present study, given that Giancaspro found a strong effect of mood selection type in a study very similar to this project.

The second within-group factor is syntactic priming. To my knowledge, no previous HL acquisition research has tested the impact of syntactic priming on HSs' grammatical variability. If HSs have acquired the functional features underlying subjunctive mood in Spanish (Chapter 3) but struggle to access those features during production (Cuza & Pérez-Tattam, 2016; Perez-Cortes, 2016; Putnam & Sánchez, 2013), it is expected that they will produce more subjunctive mood (e.g., exhibit less variability) after being exposed to subjunctive mood primes. The logic of this prediction comes from Jiang (2000), who argues that "the presence or absence of priming might provide an

indication of whether the morphological specifications have been integrated into the lexical entry" (p. 67).

The third and final within-group factor is lexical frequency. Since Dorian (1981) and Gal (1989), researchers have suggested that lexical frequency may have a significant impact on HSs' knowledge of grammatical properties. Nonetheless, I am unaware of any studies which employ experimental methodologies to examine the relationship between lexical frequency and HSs' productive and receptive knowledge of grammatical forms. There is at least one main reason to hypothesize that HSs will exhibit less variable subjunctive knowledge with more frequent verbs. Putnam & Sánchez (2013) argue that HSs, as they activate the HL less and less frequently over time, may experience difficulty accessing functional features on lexical items in the HL. If this is so, it is reasonable to expect that frequent lexical items will be less affected, therefore leading HSs to show less productive and receptive variability with subjunctive mood on frequent verbs.

1.4. Outline of the remaining chapters

The remainder of the dissertation will be organized as follows. In Chapter 2, I will present a broad perspective of HS differences, defining the notions of HS divergence and variability and, in addition, reviewing three previous accounts seeking to explain these patterns. In Chapter 3, I will turn my attention to subjunctive mood in Spanish, summarizing previous research on its syntactic structure as well as its acquisition by monolingual and bilingual native speakers.

In Chapter 4, I will review the methodology of the dissertation, introducing the participant groups as well as the experimental tasks used to gauge their knowledge of

subjunctive mood. In Chapter 5, I will present the results of the experimental tasks while in Chapter 6, I will discuss the implications of those results for our understanding of HS divergence and variability as well as our approach to HL acquisition research more generally. In Chapter 7, I will conclude the dissertation by recapping the main findings of the study and briefly discussing both (a) limitations of the present study and (b) implications for future research on HL acquisition.

CHAPTER 2:
DIVERGENCE AND VARIABILITY IN HSs

2.1 Introduction

Ever since HSs have been identified as a natural class, researchers have noted that HSs' knowledge of the heritage language (HL) often takes a different shape, quantiatively or qualitatively, from the knowledge demonstrated by dominant speakers of the HL.

These heritage speaker differences have been observed with child (e.g., Cuza, 2016; Cuza & Pérez-Tattam, 2016; Montrul & Potowski, 2007) and adult HSs across a wide range of HLs including Arabic (e.g., Albirini, Benmamoun & Saadah, 2011; Albirini & Benmamoun, 2014, 2015), Dutch (e.g. Hulsen, de Bot & Weltens, 1998), French (Kupisch, Akpinar & Stöhr, 2013), Gaelic (e.g., Dorian, 1981), German (e.g., Hopp & Putnam, 2015), Hebrew (e.g., Kaufman, 1998), Hindi (e.g., Montrul, Bhatt & Bhatia, 2012), Hungarian (e.g., Gal, 1989), Icelandic (e.g., Putnam & Arnbjornsdottir, 2015), Inuttitut (e.g., Sherkina Lieber, 2015), Korean (e.g., Lee, 2011), Norwegian (e.g., Lohndal & Westergaard, 2016), Nubian (e.g., Rouchdy, 1989), Romanian (e.g., Montrul, Bhatt & Ghirju, 2015), Russian (e.g., Polinsky, 2006, 2008, 2011, *inter alia*), Spanish (e.g., Montrul, 2002, 2008, *inter alia*), Turkish (Bayram, Rothman, Iverson, Kupisch, Miller, Puig-Mayenco & Westergaard, 2017), among many other HLs.

Not only have HS differences been observed in many HLs, they have also been documented in many domains and sub-domains of the grammar. While most research on HS differences up to this point has focused on syntax (e.g., Cuza & Frank, 2011, 2015; Polinsky, 2011; Putnam & Salmons, 2013), nominal morphology (e.g., Lohndal & Westergaard, 2016; Montrul, 2004b; Montrul, Davidson, de la Fuente & Foote, 2014; Montrul & Potowski, 2007; Polinsky, 2008) and verbal morphology (e.g., Mikhaylova,

2012; Montrul, 2007, 2009; van Osch & Sleeman, *forthcoming*; Perez-Cortes, 2016; Sherkina-Lieber, 2015), HS differences have also been observed in phonology (e.g., Kim, 2016), discourse (e.g., Sánchez-Muñoz, 2009) and other HL domains.

Many of the studies which document HS differences have been completed in the US, where HSs often receive less support for the maintenance of their HLs. It appears to be the case, moreover, that HSs in countries where bilingualism is more supported and socially accepted exhibit less differences (e.g., Kupisch et al, 2013) from dominant-speakers[2] than HSs in the US and other, more subtractive bilingual contexts[3]. Nonetheless, HS differences are not restricted to the US context, as evidenced by research from Austria (Gal, 1989), Canada (Sherkina-Lieber, 2015; Valenzuela, Iverson, Rothman, Borg, Pascual y Cabo & Pinto, 2015), Egypt (Rouchdy, 1989), and the Netherlands (Boumans, 2006; van Osch & Sleeman, *forthcoming*), among other places.

While most studies of HL acquisition do find evidence of differences between HSs and dominant-speakers of the HL, it is important to note that not *all* studies of HL acquisition reveal such differences. Leal-Méndez, Rothman & Slabakova (2015), for example, found that HSs did not diverge from Spanish-dominant speakers in their sensitivity to discourse-sensitive clitic left dislocations and focus fronting constructions in

---

[2]In this dissertation, I use "dominant speaker" as an umbrella term encompassing both (a) monolingual speakers (who are necessarily dominant in their only language) and (b) bilingual speakers with higher global dominance in one of their two languages (e.g., first generation immigrants who remain more proficient in their L1 despite spending many years in an L2 context). It is necessary to create such a term because in HL acquisition research, HSs are sometimes compared to monolingual controls (e.g., van Osch & Sleeman, forthcoming) and sometimes compared to bilingual controls (e.g., Perez-Cortes, 2016). Though these two types of controls are different, what they have in common is dominance in the first language, hence the creation of the more general term "dominant speakers." For more on dominance, see Grosjean (2008) and Silva Corvalán & Treffers-Daller (2016).

[3] This does not imply, however, that the HS experience is uniform in all countries where bilingualism is relatively more supported. HSs of languages perceived as having a lower social status, even in countries where bilingualism is very much supported, may nonetheless come to differ sharply from dominant-speakers (e.g. Boumans, 2006).

Spanish. Similarly, Kupisch, Lein, Barton, Schröder, Stangen & Stohr (2014) report that HSs of French living in Germany exhibited target-like sensitivity to French adjective placement, which differs across German and French. These two studies are by no means the only studies to find HSs performing similarly to bilingual controls.

Though *some* HSs do not diverge from dominant-speakers, *most* exhibit differences, as has been observed in many HLs, across many domains of the grammar and in many societal contexts. It seems obvious, therefore, that such differences are a fundamental, if not inevitable, part of acquiring and maintaining knowledge of a HL.

Despite the frequency with which such differences have been observed, there is still no prevailing account of (a) why these differences emerge and (b) what they reveal about the nature of HL acquisition and bilingual language acquisition more generally. What factors cause HSs to differ from dominant-speakers in their knowledge of the HL? Similarly, what factors make HSs more or less likely to differ more from dominant-speakers with some properties of the HL grammar than with other properties? Finally, and most importantly for the present study, how do we conceptualize HS grammars which differ, to differing degrees, from the grammatical systems of dominant-speakers of the HL?

The remainder of Chapter 2 will address these questions. In Section 2.2, I will discuss and operationalize what I call HS divergence (Section 2.2.1.) and HS variability (Section 2.2.2). The primary goals of this section are (a) to clarify what I mean (and do not mean) by HS differences and also (b) to illustrate how such differences present a challenge for linguistic theory. Then, in Section 2.3, I will discuss three current accounts of HS differences, highlighting strengths, weaknesses and predictions of each account.

2.2. Defining heritage speaker differences

So far in Chapter 2, I have made frequent references to so-called HS differences. Prior to exploring the nature and causes of these HS differences in Section 2.3, however, we must first define and operationalize the notion of HS differences. In the remainder of Section 2.2., therefore, I will examine HS differences from two opposing perspectives. First, in Section 2.2.1, I will examine the *between-group* perspective of HS differences, considering how HSs differ from dominant-speakers of the HL and introducing the notion of "HS divergence." Second, in Section 2.2.2, I will explore the *within-group* perspective of HS differences, addressing how HSs differ from *themselves* (diachronically and synchronically) and proposing the notion of "HS variability."

2.2.1. HS divergence

In the present study, HS *divergence* refers to those situations in which HSs exhibit linguistic behavior(s) not known to be demonstrated by dominant speakers of the HL. I have chosen to use the term divergence due to its neutral connotation, as pointed out by Scontras, Fuchs & Polinsky (2015). Nonetheless, the definition of divergence that I put forth here is my own.

Perhaps the clearest example of HS divergence can be observed with gender agreement in Spanish DPs. In Spanish, nouns are marked for gender. Both determiners (e.g., *el/la* ('the')) and adjectives (e.g., *bonito*), therefore, match the gender of the nouns that they modify. To my knowledge, there is no evidence showing that dominant speakers of Spanish (e.g., monolingual Spanish speakers or 1st generation immigrants from Spanish-

speaking countries) produce or accept gender "mismatches" in Spanish, e.g., *la_{FEM} $casa_{FEM}$ $bonito_{MASC}$, 'the beautiful house.'

Nonetheless, child and adult HSs of Spanish have been shown to (a) produce and (b) accept gender mismatches in Spanish as a HL (e.g., Cuza & Pérez-Tattam, 2016; Montrul, Foote, & Perpiñán, 2008; Montrul & Potowski, 2007). Because these HSs' production and acceptance of gender "mismatches" in Spanish has not been observed in dominant-speakers of Spanish, then I consider this linguistic behavior to be an example of HS divergence. Whatever the underlying reasons for this behavior (See Section 2.3 for three possible accounts), it is apparent that the HSs, in producing gender mismatches, are innovating beyond the input that they receive[4].

Divergence, as defined here, is not limited to HL acquisition. In fact, language acquirers innovate beyond the input in other monolingual and multilingual acquisition scenarios. In L1 acquisition, for example, children produce overgeneralized morphological forms, e.g., *goed*, which they clearly do not hear in the input that they receive from parents. Such L1 divergence can have different underlying causes. One possibility is that L1 learners producing such forms have different underlying grammatical representations (e.g., Legate & Yang, 2002; Roeper, 1999). A second possibility, argued in Pinker & Ullman (2002), is that children's production of these divergent morphological forms is attributable

[4] It is possible, and perhaps even likely, that many HSs hear gender mismatches in the Spanish input that they receive from other HSs (e.g., siblings, cousins, classmates…etc…). I do not deny that this qualitatively different input may have an influential role in HSs' production of "gender mismatched" forms. That said, it is clear that any complete account of HS divergence (with gender agreement as well as other properties) must go beyond this possibility for at least one primary reason. The HSs (e.g., only children) who first begin producing gender mismatches, unless we find evidence of such mismatches in 1st generation immigrants' command of gender, must be innovating beyond the input that they receive. That is to say, for some HSs to emulate or imitate "gender mismatch" forms, some HSs must begin producing these innovative forms, despite not having heard them in their own primary linguistic input. Similar arguments have been put forward by Sharwood Smith & van Buren (1991) and Putnam & Sánchez (2013).

to a (temporary) inability to access irregular verb forms from so-called declarative memory. Broadly speaking then, L1 divergence can be the result of either representational or lexical differences, respectively.

In L2/L3 acquisition, children and adults produce divergent forms in the L2/L3, often, though not necessarily, as the direct result of identifiable transfer from the L1. Spanish speakers, for example, produce divergent, Spanish-like differential object marking (DOM) in Brazilian Portuguese (e.g., Montrul, Dias & Santos, 2011), despite the fact that Brazilian Portuguese lacks DOM and these forms do not appear in the input provided by Brazilian Portuguese speakers. Like the monolingual English-speaking children discussed in the previous paragraph, these learners of BP innovate *beyond* the input, producing divergent forms with other root causes (e.g., L1 transfer).

HS divergence, though its causes may be more complex than in L1 or L2/L3 acquistion scenarios, is just another instantiation of language acquisition divergence more generally. In the next two paragraphs, I will consider two different dimensions of HS divergence, addressing both its time course and its relationship to L1 transfer.

Montrul (2002, 2008, *inter alia*) argues that HS divergence (in her terminology, "incomplete acquisition"; see Pascual y Cabo & Rothman, 2012; Putnam & Sánchez, 2013; and Kupisch & Rothman, *forthcoming* for criticisms of this label) can appear at two different relative points in development. The first possibility, which Montrul calls "incomplete acquisition," is that a HS demonstrating divergence with a certain property *never* came to acquire categorical knowledge of that property. Imagine, for example, the case of a five-year old HS who produces gender agreement accurately in 75% of all Spanish DPs. At age 18, after many years in an English-dominant environment, this same HS still

produces gender agreement around 75% of the time, quite unlike dominant-speakers of Spanish. This HS's divergent performance, therefore, appears to be the result of never having developed a categorical, invariant knowledge of gender agreement in Spanish.

The second possibility, which Montrul calls "attrition," is that a HS demonstrating divergence with a certain property *did reach* categorical knowledge of that property before later coming to diverge from the performance of Spanish-dominant speakers. Imagine now the case of a HS who, at age five, exhibits 100% accuracy with gender agreement in Spanish, only to perform less accurately with gender agreement (e.g., 75%) later in life. This HS, unlike the HS in the previous paragraph, diverges from Spanish-dominant speakers later in development, suggesting that his divergence is the result of having acquired, then subsequently "lost," linguistic knowledge.

It is worth noting that these two possibilities are not mutually exclusive. It is perfectly possible, for example, that a HS whose gender agreement accuracy peaks at 50% (thereby falling into the first temporal possibility of "incomplete acquisition") also experiences attrition, later producing gender agreement with even lower overall accuracy (e.g., 25%). Ideally, testing for the possibility of incomplete acquisition and/or attrition involves longitudinal data collection (Andersen, 1999), though researchers tend to employ cross-sectional designs (Polinsky, 2011).

A third account for HSs' divergence, which could theoretically apply in cases of both "incomplete acquisition" or "attrition," is that HSs who produce gender "mismatches" in Spanish maintain categorical knowledge of the relevant syntactic features (gender) and processes (agreement) but have unstable knowledge of the lexical (gender) specification of

particular nouns in Spanish. As such, their divergent gender production could be lexical, rather than syntactic, in nature.

Just as HS divergence can emerge at different points in development (e.g., before or after having reached categorical knowledge of a property), it can also emerge as a result of both (a) linguistic transfer from the societally dominant language or (b) other underlying causes. In this sense, HS divergence shares common ground with both L2/L3 divergence, which is typically the result of language transfer, and L1 divergence, which must be driven by other internal factors.

One clear example of transfer-induced HS divergence is the case of bare plural DPs in Spanish. In English, plural generic nouns are expressed with so-called "bare DPs," which lack definite articles (e.g., *bears* are strong), while plural specific nouns are expressed with DPs headed by definite articles (e.g., *the bears* are eating our lunch). In Spanish, however, both generic and specific plural nouns require DPs headed by definite articles (e.g., *los osos son fuertes*; *los osos están comiendo nuestro almuerzo*). Montrul & Ionin (2010) found that HSs of Spanish accepted bare plural nouns in Spanish, thereby diverging from the Spanish-dominant controls. Based on the fact that English allows for bare plural nouns in generic contexts, it is likely, the authors argue, that the HSs' divergence in this case is the result of transfer from English.

As highlighted by Scontras et al (2015), it is not always the case that we can identify a feature or property in the dominant language that, via transfer, leads to a certain instance of divergence in the HL. (After all, it is not possible for a speaker to transfer a property that he does not have.) When HSs diverge with HL properties that do not have a clear "equivalent" in the dominant-language, therefore, it is clear that this divergence must be

attributable to factors other than linguistic transfer. One strong example of such a case comes from HSs' production of continuous and discontinuous sentential negation in Arabic. Albirini & Benmamoun (2015) found that HSs exhibit a tendency to perform more accurately with continuous, as opposed to discontinuous, negation. Neither form has a clear equivalent in English, so the authors attribute the HSs' increased variability with discontinuous negation to the higher complexity of these forms, rather than any particular mapping or influence from the structure of English itself.

In this section, I have focused on the *between-groups* perspective of HS difference, defining and operationalizing what I call HS divergence. To summarize, I have defined HS divergence as HSs' production of linguistic forms which are not attested in the input provided to them by dominant-speakers of the HL. I have argued that such divergence, while discussed here in the context of HL acquisition, is not *unique* to HL acquisition, as both L1 and L2/L3 acquirers also exhibit clear divergence from the input that they receive. To show the breadth of the phenomenon, I have categorized HS divergence along two dimensions: time course and dominant-language transfer. With respect to time course, I followed Montrul in arguing that HS divergence can emerge either before or after a HS has reached categorical knowledge[5] of a given property. In the case of dominant-language transfer, I illustrated, using examples from different HLs, how HS divergence is sometimes, but certainly not always, attributable to dominant language transfer.

[5] By categorical knowledge, I refer simply to the hypothetical point at which a given HS produces a certain linguistic form (e.g., gender agreement in the DP) invariantly, that is to say, 100% of the time. Without testing a HS's parents (and other sources of HL input), however, it is impossible to know whether he/she was ever exposed to "categorical" knowledge of this linguistic form in the first place.

We will consider possible explanations of this HS divergence in Section 2.3. For now, however, we must first define and operationalize the *within-group* perspective on HS differences: HS variability. Section 2.2.2 addresses this topic in further detail.

2.2.2. HS variability

In Section 2.2.1, we saw that sometimes HSs "diverge" from dominant-speakers, producing and/or accepting qualitatively distinctive forms (e.g., gender mismatched DPs) in the HL. This HS divergence is defined as those situations when HSs differ from dominant-speakers of the HL.

While it is common for HSs to produce so-called "divergent" forms in the HL, it is rarely the case that they produce these forms *exclusively* (e.g., *only* producing gender mismatched DPs). A far more common outcome, based on a thorough survey of the HL acquisition literature (e.g., Albirini & Benmamoun, 2015 for negation in heritage Arabic; e.g., Montrul, Bhatt & Bhatia, 2012 for case marking in heritage Hindi; Polinsky, 2008 for gender marking in heritage Russian; Montrul, Bhatt & Ghirju, 2015 for case marking in heritage Romanian; Montrul, 2002, 2009 for verbal morphology in heritage Spanish; among many other HLs and properties) is that HSs will alternate, to differing degrees, between divergent forms and target forms, thereby exhibiting a second type of HS difference that I will call HS variability. (For a similar use of the term variability in second language acquisition, see McCarthy (2008).) Flores (2015) makes this point when she asserts that "most studies on heritage language acquisition have shown…that the speakers tend to produce certain structures in both target-like and target-deviant manners" (p. 253).

Unlike HS divergence, which refers to between-group differences, HS variability refers to the variation and differences exhibited *within* the grammatical knowledge of individual HSs. In the remainder of this section, I will define this concept of HS variability and discuss different surface forms that it can take in HL acquisition.

In the present study, HS variability refers to those situations in which a single HS produces a given linguistic form (e.g., word order, morphological inflection…etc…) variably, alternating between "target" and "divergent" forms in structures where dominant-speakers produce only "target" forms. Because not all HS divergence necessarily involves such intra-individual alternation, HS variability can be considered a subtype of the broad category of HS divergence. It is possible, though not common, for HSs to exhibit divergence without exhibiting variability, e.g., by categorically producing a non-target word order or inflection which is never produced by dominant-speakers of the HL[6]. In such a scenario, the HS acquires a grammar that is qualitatively different and yet internally invariant. When I say HS variability, I refer only to situations where individual HSs alternate between "target" and "divergent" variants of a given linguistic form.

Discussing variability, of course, invites comparisons with variationist linguistics, where researchers seek to predict variability with certain linguistic properties based on internal and external predictors, such as tense (internal) or social class (external). When I

---

[6] Some HSs of Russian appear to have developed a stable and invariant two-case system for lexical NPs, which differs dramatically from the six-case system used by dominant speakers of Russian (e.g., first generation immigrants). Even if these HSs' use of the two-case system is shown to be highly systematic, as it likely is, the HSs' knowledge of nominal morphology remains "divergent," in the sense that it (almost certainly) differs sharply from the input that they received in Russian from parents and other primary sources of input. We do not know for sure whether these HSs learned a two-case nominal system from their parents, as the parents of the HSs reported in Polinsky (2006) were not tested. Nonetheless, it is exceedingly unlikely, based on ample L1 attrition research (e.g., Schmid, 2012), that the parents of these Russian HSs immigrated to the US with a six-case distinction and then, after significant time in the US, ended up using an exclusively two-case nominal system in Russian. As such, it is almost certainly the case that the HSs' two case system is an innovation beyond their home HL input.

use the term HS variability, I share variationists' enthusiasm for understanding these factors, e.g., how a Spanish-English bilingual's generational status might impact his/her likelihood of using overt, as opposed to null, subject pronouns. What separates my usage of the term variability is the type of linguistic property that I explore. While variationist linguists seek out naturally, non-categorical properties without definitive "target" forms (e.g., subject pronoun usage), I explore properties which are produced categorically and invariantly in the speech of Spanish-dominant speakers, focusing on how individual HSs alternate between mirroring, and diverging from, Spanish-dominant speakers. (For a variationist study of "obligatory" grammatical forms, see Martillo-Viner, 2017).

It is useful, once again, to further exemplify the notion of HS variability by considering the case of gender agreement in Spanish. For dominant-speakers of Spanish, gender agreement is an invariant and obligatory process. For many HSs of Spanish, however, gender agreement is produced and comprehended far more inconsistently.

Montrul, Foote & Perpiñán (2008) tested HSs' knowledge of gender agreement in a series of three experiments and found considerable variability in each. In the Oral Picture Description Task, participants saw pictures (e.g., a banana) and were asked to describe what they saw using the carrier phrase, *yo veo un(a)* ('I see a _____') followed by a noun (e.g., *plátano*) and a modifier (e.g., *maduro*). With determiners, participants' productive accuracy with gender agreement ranged from 73.4% with feminine, non-canonical nouns (e.g., *la mano*) to 100.0% with canonical, masculine, inanimate nouns (e.g., *el libro*). With adjectives, participants' productive accuracy with gender agreement ranged from 70.8% with feminine, non-canonical nouns to 99.4% with canonical, masculine, animate nouns

(e.g., *el chico*). Although the HSs' overall accuracy in this task was 89.7%, only 16 of 67 HSs (24%) performed categorically like the control group[7].

These results, when considered alongside similarly variable results in the other two experimental tasks, paint a clear picture of HS variability. While *most* HSs' produce gender agreement *most* of the time, many variably produce and accept gender mismatched forms. In other words, most HSs exhibited a strong tendency to produce and accept gender agreement while still occasionally producing and accepting gender mismatched DPs. This pattern of frequently producing gender agreement, and intermittently producing gender-mismatched DPs, is a clear example of the concept of HS variability outlined in this section.

Broadly speaking, any HS who alternates between target and divergent variants of a given property exhibits HS variability. Nonetheless, HSs' productive variability can surface in a few qualitatively different forms, a few of which I will consider here.

One possible shape of HS variability is the pattern of *weak variability*, in which HSs produce one linguistic pattern almost all of the time (e.g., 90%) and another, slightly different variation only occasionally (e.g., 10%). Weak variability, as a concept, exists independently of between-group comparisons with dominant-speakers of the HL. Nonetheless, it is easy to classify weak variability into two different subtypes based on the resemblance between the grammatical performance of HSs and dominant-speakers. On one hand, HSs may exhibit weak variability by performing almost *identically* to dominant-speakers of the HL, e.g., by producing gender agreement between determiners and nouns

[7] It is possible that these HSs' variability with gender agreement is simply an artifact of lexical gender *assignment* (specifically with non-canonical nouns), rather than syntactic *agreement*. Nonetheless, there is evidence from other HL acquisition research (e.g., Cuza & Pérez-Tattam, 2016; Montrul & Potowski, 2007) that HSs' variability with gender agreement is not always a function of different lexical gender *assignment*.

between 90 and 100% of the time (Montrul et al, 2008). On the other hand, HSs may exhibit weak variability by almost always performing *differently* from dominant-speakers of the HL. An example of this would be a HS who almost exclusively uses masculine morphology on adjectives modifying feminine nouns (e.g., la casa$_{FEM}$ bonito$_{MASC}$), yet occasionally uses "target-like" feminine morphology on such adjectives (e.g., la casa$_{FEM}$ bonita$_{FEM}$). In principle, both this hypothetical HS and a HS who more closely matches the patterns of Spanish-dominant speakers exemplify the notion of weak variability.

A second possible shape of HS variability is the pattern of *strong variability*, in which HSs alternate, with almost equal frequency, between two variations of a given linguistic property. An example of such *strong variability* comes from the spontaneous oral production of subjunctive mood by low proficiency HSs in Montrul (2009). Montrul reports that the low proficiency HSs in this study produce subjunctive mood morphology in 44 of 110 (40%) expected subjunctive contexts (e.g., contexts in which subjunctive mood is "obligatory" for dominant-speakers). These low-proficiency HSs can be said to exhibit strong variability because they alternate almost equally between two forms, indicative (60%) and subjunctive (40%) mood morphology, in contexts where dominant speakers of Spanish use a single grammatical form.

Thus far, this discussion of HS variability has, for the sake of simplicity, focused exclusively on HSs' variability in *production*. It would be remiss, however, to exclude comprehension from this discussion of HS variability. Conceivably, HSs demonstrating each of the aforementioned variability types (e.g., weak or strong variability) with a given property might still exhibit differing levels of sensitivity to this property in *comprehension*. (After all, even HSs who do not produce the HL at all have been shown to be sensitive to

certain morphological distinctions in the HL; see Sherkina Lieber (2015).) That said, it is not in the scope of this section to address all of the many possible ways in which productive and receptive sensitivity might interact in the minds of HSs. Much more important is highlighting the foremost theoretical challenge raised by the pattern of HS variability, broadly defined: What does a HS know about a property when he/she produces it variably?

Section 2.3. will compare and contrast three different approaches to this question.

2.3. Accounts of HS divergence/variability

Undoubtedly, the most challenging element of HS research is understanding and explaining the patterns of HS divergence and HS variability that arise, to differing extents, in most studies of HL acquisition. (See Leal-Méndez, Rothman & Slabakova, 2015 for an exception, and Kupisch & Rothman, *forthcoming* for a discussion of several studies in which HSs do not diverge considerably from controls.) In seeking to explain these patterns, researchers have proposed a range of different ideas which can be grouped into three broad, conceptual classes: (i) the Input Quality approach (e.g., Pires & Rothman, 2009), which attributes divergence to emerging differences in the input that HSs receive; (ii) the Representational Differences approach (e.g., Montrul, 2002, 2008, *inter alia*), which attributes divergence to HSs' underlyingly different mental representations of linguistic features; and (iii) the Activation/Lexicalist approach (Putnam & Sánchez, 2013), which attributes divergence to HSs' reduced usage of and access to features in the HL grammar.

In Sections 2.3.1 through 2.3.3, 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 these three approaches, though presented separately, are neither

mutually exclusive, nor clearly "correct" or "incorrect". It is quite likely, in fact, that explaining and understanding certain patterns of HL acquisition data will require, at least to some extent, explicit reference to all three of these approaches.

2.3.1. The Input Quality approach

It's possible that HS divergence and variability, at least in some cases, are rooted in input quality[8]. In other words, HSs' production of seemingly divergent forms may be triggered in part by differences in the input that they receive from their parents, whose own Spanish systems may be undergoing gradual changes. It may be the case, for example, that a HS who produces occasional gender-mismatched DPs does so because he hears these forms in the input he receives.

Given my own definition of HS divergence, which assumes that HSs' exhibited grammatical "differences" can only be considered "divergent" if they are *not observed* in the input that they receive, this approach amounts to saying that *some* of what appears to be "HS divergence" is not *true* divergence but rather the acquisition, emulation, or amplification of input variability.

There are at least three different ways in which input quality can lead HSs to exhibit what appears to be qualitatively distinct grammatical outcomes. The first, and most extreme case, is when HSs fail to exhibit knowledge of a given property because that property *does not appear at all* in the input that HSs receive. The clearest example from the literature is the case of inflected infinitives in HSs of Brazilian Portuguese, presented

[8] When I use the term "input quality," I do not mean to imply that variable HL input is better or worse, in any evaluative sense, than invariant HL input. That said, I refrain from using a more neutral term, such as "input inconsistency," because input quality is an established and easily recognizable term in HL acquisition research.

by Pires & Rothman (2009). Inflected infinitives, though still a part of highly formal Brazilian Portuguese registers, are used rarely, if ever, in the informal, colloquial speech of Brazilian Portuguese speakers. Not surprisingly then, HSs of Brazilian Portuguese in the US, who lack formal exposure to BP, neither produce nor understand inflected infinitives. If compared to educated BP speakers, who learned and mastered inflected infinitives in school, these HSs seem to exhibit "divergent" linguistic behavior. In light of the fact that they didn't *hear* these forms, however, their seemingly divergent linguistic behavior really amounts to what Pires & Rothman have called "missing input competence divergence." In other words, HSs differ when differences in input quality (e.g., what forms hear and do not hear) lead to insufficient opportunities to acquire certain, higher-register linguistic forms.

It's important to note that the case of inflected infinitives, though conceptually and methodologically revealing, is anomalous, given that almost all linguistic properties tested in HL acquisition research are abundant in the input received by HSs. There is no doubt, for example, that HSs of Spanish are exposed to certain, commonly-researched grammatical forms such as gender agreement and verbal morphology, which appear in nearly every Spanish utterance.

Nonetheless, even in cases where HSs *do* receive significant exposure to tested linguistic properties, it is possible for input quality to play a prominent role in shaping the apparent divergence of their HL grammars. If first-generation immigrants, who provide second generation HSs with significant input in the HL, begin to exhibit emerging variability with certain properties of the HL, it is possible that HSs will either *replicate* or *amplify* this variability as they form their own grammatical systems in the HL. I consider both these possibilities in the following paragraphs.

Montrul & Sánchez Walker (2013) found evidence suggesting that HSs, when exposed to emerging variability from Spanish-dominant speakers, replicate this variability in their own HL production. In their study of differential object marking (DOM), Montrul & Sánchez Walker compared HSs' production of DOM with the DOM production of first-generation immigrants. Crucially, the first-generation immigrants (average time in the US: 25.9 years) were similar in profile to the parents of the HSs, making it likely that the characteristics of their Spanish are comparable to what the HSs in the study heard from their parents at home. In a picture description task, both the simultaneous HSs and the sequential HSs exhibited substantial variability with DOM, producing it only 77.7% and 77.0%, respectively, with [+animate, +specific] direct objects, where DOM is expected to be obligatory. The results from the first-generation immigrants, however, indicate that the HSs' variability with DOM is just a simple replication of variability in the input that they receive. The first-generation immigrants in the study also exhibited nearly identical DOM variability, producing it 81.3% of the time with [+animate, +specific] direct objects. It is reasonable to conclude, therefore, that the HSs' variable performance with DOM is not HS divergence but a simple replication of emerging variability in the input that they receive[9].

Thus far, we have seen an example of HSs hearing, and then emulating, input variability. It is also possible, however, that HSs will respond differently to the presence of input variability, amplifying it into qualitatively distinct grammars rather than simply reproducing its frequencies.

---

[9] Another interpretation of this finding is that the original linguistic assumption, e.g., that animate, specific direct objects are obligatorily marked with DOM, is not descriptively accurate for first generation immigrants or monolingual speakers. However, Montrul & Sánchez-Walker (2013) also tested monolingual children and adults in Mexico and found that both groups produce DOM over 95% of the time with animate, specific direct objects.

One example of this possibility comes from Pascual y Cabo (2013), who examined HSs' knowledge of the reverse psychological predicate, *gustar*. The verb *gustar*, which is a so-called Class III verb, is non-agentive and therefore, does not typically appear in passive constructions (**La pizza fue gustada por Pau*). Class II psychological predicates like *molestar*, on the other hand, *are* agentive and therefore can be passivized (*El gato fue molestado por el perro*). In an acceptability judgment task, the adult HSs in Pascual y Cabo's study showed a distinct tendency to accept *gustar* in passive constructions, suggesting that they accept, or are beginning to accept, *gustar* as a Class II predicate.

What causes the HSs to exhibit this emerging tendency to reanalyze or reclassify *gustar*? Pascual y Cabo argues that the HSs' qualitatively distinct grammatical performance with *gustar* is driven by emerging *gustar* instability in the grammatical systems of first generation immigrants. It is important to clarify that the first generation immigrants in the study do not show much of a tendency to accept *gustar* in passive constructions. However, Pascual y Cabo posits that even minor changes and variability in the first-generation immigrants' use of *gustar* (e.g., de Prada & Pascual y Cabo, 2011) could be enough to lead HSs towards a qualitatively distinct analysis in which *gustar* can act as both a Class III or a Class II predicate, respectively[10]. In other words, the presence of qualitatively different input may lead HSs' to posit divergent grammatical analyses, perhaps for reasons of economy (Scontras et al, 2015).

---

[10] It is also possible that HSs' tendency to treat *gustar* as a Class II predicate is at least partially the consequence of lexical transfer from English, where the semantically equivalent predicate *like* is a Class II predicate which can be passivized (e.g., *The cake was liked by the guests*). That is to say, the instability mentioned above could be triggered by lexical influence from English *as well as* exposure to increasingly variable uses of *gustar* in the Spanish input.

A second example of HSs amplifying, rather than simply replicating, incipient changes in the input comes from Bookhamer's (2013) analysis of subjunctive mood in US Spanish. Using a Spanish corpus made up of interviews with first- and second-generation Spanish speakers living in New York City, Bookhamer finds evidence of emerging variability in first-generation speakers' usage of subjunctive mood in so-called indefinite contexts (e.g., *busco uno que venga*). The first generation speakers in his study use subjunctive only 84% of the time in these constructions, exhibiting an emerging tendency to produce indicative mood forms instead.

How do the second-generation HSs respond to this qualitatively distinct input? Given that they only produce subjunctive 64% of the time in these contexts, e.g., 20% less than the first-generation speakers, it seems apparent that the second-generation HSs amplify the incipient variability that they receive in the input. In a recent discussion of this data set, Otheguy (2014) makes a similar point, arguing that the grammar of the second-generation HSs differs from that of the first-generation "in having moved a few notches the line between obligatory and variable environments, as well as in *having yielded more space to the indicative in those environments where previous generations had already been making room for it*" (p. 306, emphasis mine). Emerging variability, therefore, is not just replicated but also *expanded* by second-generation HSs.

The Input Quality approach (e.g., Pires & Rothman, 2009) is critical in HL acquisition research because it reminds us of the importance of taking into consideration *what HSs hear* in the HL input. Sometimes, what appear to be divergent characteristics of HSs' grammatical production are, in fact, directly traceable to differences that HSs have heard from first-generation speakers, whose own grammatical systems are in flux due to

contact with English and/or reduced use of the L1. Acknowledging and accounting for this possibility has become imperative in HL research and, unquestionably, has made HL acquisition research stronger.

Nonetheless, there are limitations to the Input Quality approach, which make clear the necessity of proposing additional accounts of HS divergence and variability. The first clear limitation is that HS divergence is observed even with grammatical properties which (a) are abundant in the input (e.g., Montrul, 2016b) and (b) do *not* exhibit emerging variability in the speech of first-generation immigrants. As Montrul, Bhatt & Ghirju (2015) argue, "what sets HSs apart is…eventual mastery of their heritage language in early adulthood can vary significantly from the end-state achieved by their own parents and from other native speakers of the same language in their linguistic community of origin" (p. 565). In Section 2.2.1., we discussed this phenomenon, pointing out that HSs sometimes produce and accept gender-mismatched Spanish DPs, despite the fact that no study has found such divergence in first-generation immigrants. It must be the case, therefore, that HSs' divergence with grammatical gender in Spanish (as well as other similarly invariant properties, e.g., preposition stranding: Pascual y Cabo & Gómez Soler, 2015) is driven by factors other than input quality, as has been openly acknowledged by proponents of this approach (e.g., Pascual y Cabo & Rothman, 2012).

A second limitation of the Input Quality approach is that it does not address (nor does it intend to address) the underlying nature of HS variability, that is to say, the *within-group* perspective of HS differences. Consider, for example, the case of DOM in Spanish. If HSs produce DOM approximately 80% of the time, due to similarly variable output from the first-generation, what underlying grammatical knowledge do they have of DOM? Does

their variability imply that their knowledge of DOM is probabilistic? Or is it the case that these HSs have target-like knowledge of DOM which they find increasingly difficult to access and reliably produce in real time? With respect to these important questions, the Input Quality approach is decidedly and openly agnostic. Consequently, researchers must explore explanations beyond input quality to understand HS divergence and variability.

In Sections 2.3.2. and 2.3.3., respectively, I will abstract away from input quality differences and consider what it means, from the standpoint of grammatical knowledge, for HSs to exhibit (a) HS divergence and (b) HS variability with linguistic properties.

2.3.2. The Representational Differences approach

The intuition behind what I call the Representational Differences approach (Montrul, 2002, 2008, *inter alia*) is as follows: when HSs differ from dominant-speakers, e.g., by exhibiting divergent variability with verbal morphology, it is because they have acquired, or ended up with, different mental representations of the properties/features which underlie these particular linguistic forms. From this perspective, differences between HSs and dominant-speakers imply differences in underlying knowledge.

The first, and most prominent, example of this approach comes from Montrul's seminal (2002) paper on Spanish HSs' knowledge of aspectual morphology.

In the first part of the experiment, HSs were tested on their ability to produce, both orally and in writing, perfective and imperfective past-tense morphology in Spanish. In the oral production task, the Spanish-dominant controls displayed categorical knowledge of aspectual morphology, producing both perfective and imperfective forms with 100% accuracy. In the written production task, which asked participants to select between

perfective and imperfective forms within the context of a letter, the controls again performed almost invariantly, reaching 96.5% accuracy. The HS groups, on the other hand, exhibited more variability than the controls in both oral production, where their average accuracy ranged from 77.14% to 100.0%, and written production, where they responded accurately between 85.0% and 91.3% of the time, respectively.

In the second part of the study, Montrul tested HSs' interpretation of aspectual morphology, employing both a Sentence Conjunction Task and a Truth Value Judgment Task. As in the productive tasks, the Spanish-dominant controls performed almost categorically, demonstrating clear sensitivity to the semantic distinctions between perfective and imperfective forms in Spanish. The HSs, once again, exhibited more variability than the controls in both interpretive tasks, drawing less of a distinction between imperfective and perfective forms, respectively.

On the basis of these productive *and* interpretive between-group differences, Montrul posits representational differences between the morphological knowledge of the two groups, stating that "at least in the area of aspectual semantic interpretations, many bilinguals *never* converge on the Spanish system of native speakers" (p. 58). Unlike Spanish-dominant controls, who have "completely" acquired these morphological distinctions, HSs exhibit variability due to underlyingly "incomplete" knowledge of this area of the Spanish grammar. Regardless of the negative sociopolitical connotations of the term "incomplete" (e.g., Kupisch & Rothman, *forthcoming*; Otheguy, 2014; Pascual y Cabo & Rothman, 2012; Putnam & Sánchez, 2013, *inter alia*), it is not difficult to see why Montrul has employed this term to describe HS grammars that exhibit variability with properties that do not appear to be variable for dominant speakers of the HL.

Nonetheless, there are significant empirical and conceptual difficulties with the Representational Differences approach, which seems to attribute any between-group differences, no matter how small, to differences in HSs' underlying linguistic representations. If categorical (e.g., 100%) production/interpretation is evidence of target-like knowledge and anything more variable is evidence of HSs' representational differences, then it becomes challenging to adequately capture the many and varied ways in which HSs differ from dominant-speakers. To illustrate this point, let's return to some of the HS data presented in Montrul (2002).

On the Sentence Conjunction Task, which tested HSs' ability to detect semantic irregularities associated with the infelicitous use of aspectual morphology, two simultaneous HSs failed to distinguish between perfective and imperfective morphology in *any* of the three experimental conditions (achievements, accomplishments and states). These two HSs clearly differ from the controls, almost all of whom (17 of 20) significantly distinguished between the two aspectual forms in all three experimental conditions[11]. Because these two HSs do not appear to comprehend differences between perfective and imperfective morphology, regardless of predicate type, it seems safe to assume that they have arrived at different representations of aspectual morphology in Spanish. In this instance at least, divergence *does* imply difference.

More common, however, are HSs who diverge from controls, yet still appear to maintain at least residual sensitivity to the underlying grammatical property at hand. This pattern, which emerges in nearly every study of HL acquisition (including every study

[11] The three controls who did not perform at ceiling distinguished between perfective and imperfective morphology in two of the three relevant experimental conditions, thereby exhibiting systematic knowledge of these inflections.

listed in the Introduction of Section 2.1), defies dichotomous classification (e.g., acquired or not acquired; complete or incomplete knowledge) and problematizes the Representational Differences approach.

If HSs who differ *quantitatively* from dominant-speakers necessarily have different featural knowledge, then how come so many HSs still make the same *qualitative* grammatical distinctions as controls in production and/or comprehension? In other words, if HSs do not have the same underlying grammatical compass as dominant-speakers, then how do they so regularly arrive within close proximity of the dominant-speakers' grammatical destinations?

One reasonable possibility is that HSs have the same underlying featural knowledge as the controls but occasionally diverge from them when they fail to "map the semantic features onto morphophonological material" (p. 51). In other words, it may be the case that HS divergence is located in the lexicon, specifically in the mapping of functional features (FFs) to individual lexical items. Montrul (2002) raises, but then rules out this possibility, due to between-group differences between the HSs and the controls on the two receptive tasks of the study, which she interprets as evidence that HSs' underlying "knowledge (mental representation) of the functional category AspP is either incomplete or lacking altogether" (p. 51). This account, which again assumes that between-group differences between HSs and controls signify underlying differences in the two groups' featural knowledge, faces a couple of significant explanatory obstacles.

First, as stated above, it remains unclear how HSs with underlying representational differences could perform above chance in the production and comprehension of aspectual morphology, across three different experimental tasks. Second, and on a very closely

related note, it is unclear what the term "incomplete" (or representational difference) really means in the case of these divergent HSs. Aside from the possibility that HSs' "incompleteness" is a function of mapping problems, which Montrul (2002) rules out, it's not obvious what exactly "incomplete" knowledge of aspectual (or any other type of) morphology might mean. Does it mean that HSs have acquired the same functional features as controls but in a less stable way? Or does it mean that their featural knowledge itself is somehow representationally distinct? The answers to these critical questions remain unclear both for this approach and the field more generally.

Montrul (2016a), following Berman (2004), has recently suggested new terminology to discuss the non-binary nature of HSs' linguistic outcomes. Specifically, she draws a distinction between what she calls *acquisition* and *mastery*. For Montrul, acquisition is when speakers "show productive use of a given form" though do not necessarily "use it in all obligatory contexts" (p. 123). Mastery, on the other hand, implies that speakers have reached categorical performance with a given form, such that it is "always produced with 90-100% accuracy" (p. 123). In other words, acquisition implies identifiable, though not necessarily categorical, linguistic knowledge of a property while mastery of that property implies knowledge that is clearly categorical.

This welcome terminological distinction opens the door for more fine-grained (e.g., non-binary) conversation about HSs and the ways in which they exhibit divergence and variability in the HL. Nonetheless, the specific conceptual implications of this distinction remain as yet undefined. What is the difference, from the standpoint of representation and/or processing, between a HS who has *acquired* aspectual morphology (e.g., producing 75% target forms) and a HS who has *mastered* it (e.g., producing 100% target forms)? In

the absence of specific answers to this question, the acquisition/mastery distinction is no more than a useful terminological tool for separating HSs performing variably from those performing categorically.

2.3.3. Activation/Lexicalist approach

Thus far, we have looked at two possible explanations for HS divergence and variability. The first possibility is the Input Quality approach, which attributes *some* HS divergence to the qualitatively different HL input to which HSs are sometimes exposed. While this approach accounts for some patterns of HS divergence, it fails to explain why HSs exhibit divergence even when they are *not* exposed to qualitatively different input (e.g., with grammatical gender in Spanish). Additionally, the Input Quality approach offers no explanation for HS variability. If a HS exhibits variability, even if it *can* be traced back to input quality differences, how do we model such intra-speaker inconsistency in the bilingual mind? This approach is silent on this question.

The second possibility is the Representational Differences approach, which attributes HS divergence to differences in HSs' underlying grammatical representations. Assuming representational differences is, of course, a straightforward way to account for both (a) how HSs diverge from dominant-speakers of the HL and (b) how HSs exhibit intra-speaker variability. Nonetheless, this approach fails to provide a compelling story for a common, but particularly challenging puzzle of HL acquisition, namely, how HSs with variable HL grammars can make the same qualitative grammatical distinctions as dominant-speakers of the HL. It is apparent, therefore, that this dichotomous approach to

HL differences lacks the flexibility to account for HS grammars which at once resemble, *and* differ from, the grammars of dominant speakers.

In this section, I will introduce the Activation/Lexicalist approach, which addresses some of the shortcomings of the Input Quality and Representational Differences approaches, respectively. The chief advantage of this approach is its flexibility and versatility in accounting for the complex patterns that characterize HL acquisition and maintenance. Not only does this approach provide a story for how HSs exhibit divergence—even without being exposed to differences in input quality—it also explains how HSs with underlyingly target-like knowledge of a grammatical property might still exhibit variability in the production of that property. Finally, by taking into consideration the role of the HL lexicon in HS divergence and variability, the Activation/Lexicalist approach stimulates new lines of inquiry in HL acquisition research.

Putnam & Sánchez (2013) assume that HSs' grammatical knowledge of the HL is not an outcome or end-state but a dynamic system which fluctuates over time according to the frequency with which HSs activate the HL for production and comprehension purposes. These authors are not the only researchers to point out that HSs' usage of the HL fluctuates over time (e.g., Silva Corvalán, 1994, among many others). What's unique about Putnam & Sánchez (2013) is that they attempt to model *how* such shifts in HL activation over the lifetime impact HSs' expressed knowledge of the HL, resulting in some of the hallmark findings of HL acquisition. Of particular interest to Putnam & Sánchez are HSs' production/comprehension asymmetries, which the authors attribute to the increased difficulty of activating the HL for production purposes, and dominant-language transfer,

which they attribute to the reassembly of L2 functional features onto L1 lexical items. In both cases, it is the relative activation of the HL that drives HSs' linguistic differences.

To illustrate how Putnam & Sánchez (2013) can be applied to HS divergence and HS variability, as defined in the present dissertation, let's imagine the case of a HS who produces the "obligatory" verbal morpheme X just 70% of the time, unlike first-generation speakers, who produce the morpheme X categorically (100%). Because of the invariant nature of the input that he receives from first-generation speakers, this HS's variable production of morpheme X cannot be a consequence of input quality differences. How would the Activation/Lexicalist approach explain this HS's variable, and apparently innovative, production of morpheme X?

According to Putnam & Sánchez, HSs who use the HL less and less over time may come to experience "a decline in the availability of FF's [functional features]" (p. 482), which may surface in variable production of certain morphosyntactic forms. Therefore, the HS's variable production of morpheme X could be the result of having failed to access, due to reduced activation of the HL, the underlying functional feature(s) which trigger X.

Crucially, this conceptualization of HS variability is not necessarily tied to differences in the input-quality received by HSs. HSs' production of innovative and divergent forms, consequently, can occur even if their HL "input does not necessarily diminish in quality" (p. 482). Variability, in this account, is born not in the input but in the *mind* of HSs, whose relative use of the HL at any given time makes it easier or more difficult to access and produce certain forms. Usage differences, therefore, could conceivably lead different HSs with the exact same input quality to differ dramatically from one another in productive variability. While HSs who more frequently use the HL

will exhibit lower productive variability with a given form, HSs with much lower HL usage (e.g., "passive bilinguals") may exhibit *no* productive knowledge of that form at all, despite systematically comprehending it in others' speech (e.g., Sherkina-Lieber, 2015).

If HS variability emerges, at least in part, due to reduced usage of the HL, then it is no longer necessary to argue that HSs who exhibit variability with a given property lack a target-like underlying representation of that property. Consider, again, the case of the HS who produces morpheme X in 70% of all expected contexts and, say, 5% in unexpected contexts. Comparing this HS's grammatical performance to that of dominant-speakers of the HL, of course, will highlight the ways in which his grammar diverges from theirs. Analyzing his own performance across conditions (e.g., expected vs. unexpected contexts for morpheme X), however, will underscore the qualitatively target-like nature of his grammar (Bley-Vroman, 1982). Just like the dominant-speakers, this HS is more likely to produce morpheme X when it is expected (70%) than when it is not expected (5%), suggesting that his variable production is guided by the same linguistic representations that guide the dominant-speakers' invariant linguistic productions. (If his knowledge was incomplete or random, why would he demonstrate these clear distinctions?)

It is not sufficient, however, to acknowledge the target-like directionality of this HS's performance without more fully addressing his variability in production. Why, if he has the same underlying representation of the property as dominant speakers, would he produce it so much less frequently? Is it really feasible to suggest that he has acquired the functional features which trigger morpheme X but fails to access these features, due to reduced HL usage, 30% of the time? In order to strengthen the case for this possibility, I

next argue that HSs' expressed knowledge of certain properties and functional features is shaped to some extent by lexical knowledge.

It is logically possible that a HS, like the hypothetical HS in the previous section, could acquire the functional features which underlie a given form (e.g., morpheme X) and yet either (a) not always access those functional features during production (as argued so far) and/or (b) not know the ways in which that particular feature is instantiated with certain lexical items or verbal roots. Both of these logical possibilities, attested in HL acquisition research, follow from Putnam & Sánchez's assumption that the "acquisition of the lexicon and acquisition of features are distinct yet interconnected" (p. 489).

The first possibility is consistent with findings from psycholinguistic investigation of the bilingual mental lexicon. Gollan, Montoya, Cera & Sandoval (2008) argue that bilinguals, including but not limited to HSs, have weaker links between individual lexical items and their phonological and semantic characteristics. The strength of these links, however, is dependent on the relative frequency with which a speaker uses or activates his languages. As Gollan et al state, a bilingual's increased use of one of his languages "leads to improved lexical accessibility" (p. 788). It's not hard to imagine that the converse of this statement is also true; HSs who decrease their relative use of the HL may experience increased difficulty in lexical accessibility in the HL.

It is important to note that Gollan et al explored lexical accessibility and its relationship to language usage via picture-naming tasks, which limits the direct applicability of these findings to Putnam & Sánchez's analysis. Bilinguals' increased difficulty in naming pictures presumably stems from lower activation of *entire lexical items*, rather than any particular functional features associated with those lexical items

(e.g., Mood or Aspect). Nonetheless, there is additional research which supports the possibility that bilingual speakers may, for many reasons, use lexical items without necessarily accessing, or having acquired, all of their associated functional features.

Jiang (2000), in his discussion of L2 lexical knowledge, suggests that L2 learners pass through different stages of lexical knowledge in which they may exhibit differential knowledge of a word's associated syntactic and morphological features, respectively. (Though Jiang is discussing L2 acquisition, it is not difficult to see how his approach to lexical acquisition might work quite similarly in the case of HSs acquiring or maintaining HL lexical items.) At earlier stages, Jiang suggests that L2 learners can acquire the phonological/morphological forms of an L2 word without having acquired that word's syntactic and semantic features. In the final, "lexical integration stage," L2 learners' lexical representations include not just phonological and morphological features (lexemes) but also their syntactic and semantic features (lemmas). Even after reaching this stage, however, Jiang argues that L2 learners may still have difficulty accessing or retrieving some of the features associated with a lexical item during on-line language production.

What morphology, if any, will bilinguals (L2 learners or HSs) produce when they fail to access functional features during real-time language production? According to McCarthy (2008, 2012), who works within the framework of Distributed Morphology (Halle & Marantz, 1993) bilingual morphological production errors are likely to involve the use of default morphemes (e.g., masculine gender marking in Spanish) in place of more specified morphemes (e.g., feminine gender marking in Spanish), but *not* more specified morphemes in place of default morphemes. If we extend this analysis to HSs, then HSs

who temporarily fail to access functional features in on-line language production would also be expected to "overproduce" default morphological forms.

A second logical possibility is that a HS could acquire a given functional feature, say, Mood, and yet not know how that particular feature is instantiated morphologically on all lexical items. Berman (2007) argues that acquisition of vocabulary, broadly defined, is "the most salient and easily measurable facet of school-age language development" (p. 348). Part of vocabulary development, without question, involves learning the morphological inflections that go with particular lexical roots. In both monolingual and bilingual contexts, therefore, children acquiring morphologically-rich languages like Spanish must learn not just the meanings of new lexical items (e.g., *estremecer*) but also how the functional features of Spanish are instantiated on those new lexical forms (e.g., *estremezca*$_{\text{SUBJ}}$). Lardiere (2005) calls this knowledge of "which forms 'go with' which features" (p. 179) "morphological competence," which, crucially, is at least partially independent of featural knowledge. As a consequence, it is possible that HSs, especially with lower literacy and reduced exposure to the HL, will acquire a functional feature and yet not know how that feature manifests itself on certain, less frequent lexical items[12].

There are a number of previous studies which point to the HL lexicon as a potential source of HSs' demonstrated variability in grammatical knowledge. Dorian (1981), in her seminal study of Gaelic language maintenance in Scotland, studied HSs' ("semi-speakers"

[12] Once a given FF is acquired, HSs (and other speakers) must learn how that feature is instantiated on certain lexical items. HSs with less schooling in the HL will presumably have less opportunities to learn how that particular FF is instantiated on less common verbs which may not appear as much in the HL input that they receive at home. To clarify, the prediction here is that HSs will first learn how FF's are instantiated on frequent lexical items before eventually learning their instantiations on less frequent lexical items. I am not, by any means, making predictions about whether certain, more frequent FF's (however that might be quantified) are acquired before other FF's.

in her book) use of verbal morphology and found that these speakers were most likely to employ conditional morphology with frequent verbs. Similarly, Gal (1989) found that Hungarian HSs living in Austria only used causative morphology with frequent verbs.

Both of these sets of HSs, based on their oral production, appear to have (some) knowledge of the functional features that underlie conditional and causative morphology, respectively. The fact that they almost exclusively use these forms with frequent verbs, however, suggests that their knowledge of these functional features is shaped by lexical experience in the two ways discussed above and reiterated here. Either the HSs *have* acquired the underlying functional features and are only able to reliably access them with frequently activated lexical items, or they have acquired these functional features but have not learned their particular instantiations on certain, less frequent verb forms. Therefore, as Gal postulates, "it is not their [HSs'] combinatorial ability that is different but their repertoire of verb roots" (p. 327). Variability, it appears, starts in the HL lexicon.

Other, more recent studies of HL acquisition have also found that HSs' expressed knowledge of grammatical properties is closely tied to the HL lexicon. Montrul, Davidson, de la Fuente & Foote (2014) tested HSs' knowledge of gender agreement with canonical and non-canonical nouns in Spanish. In three different and differentially explicit experimental tasks evaluating gender assignment and agreement, the HSs exhibited variable, yet well-above chance accuracy. In the Gender Matching Task, the HSs' accuracy rate was 82.5% overall, though this accuracy dropped to just 74.0% with non-canonical nouns. In the Grammaticality Judgment Task, the HSs' responded accurately once again (84.4%), though their accuracy dropped with non-canonical nouns. Finally, in the Word Repetition Task, HSs were faster to repeat canonical, as opposed to non-canonical nouns.

Citing Gollan et al (2008), the authors suggest that the HSs' emerging variability with gender agreement in Spanish is attributable to weaker links between HL lexical items and their underlying gender features. Specifically, they argue that these weaker links lead to "gender assignment errors, slower retrieval of nouns in the lexicon, slower insertion of nouns in the syntax, and slower speed at computing syntactic dependencies (concord with determiners, nouns and adjectives)" (p. 135). According to the authors, all of these consequences of weaker lexical links ultimately "lead to gender agreement errors in Spanish heritage speakers." Because HSs performed well-above chance with gender-agreement in all three experimental tasks, it is appropriate to assume that the HSs have underlying knowledge of gender (and gender agreement) but exhibit variability due to differences in their representation and/or access to lexical knowledge in the HL.

The idea that HS variability is at least partially rooted in the HL lexicon is not limited to these studies. Bayram et al (2017), Kupisch et al (2013), Kupisch & van de Weijer (2016), and many others have made similar claims with different HLs and linguistic properties. (For a similar argument with L2 grammatical variability, see Kirova, 2016.)

As stated at the beginning of Section 2.3.3, exploring the role of the HL lexicon in HS variability opens up a number of fruitful lines of future research, two of which I will briefly highlight here. It is possible, as recently articulated by Lohndal & Westergaard (2016), that HSs will exhibit less variability with grammatical features that can be stored as single units alongside lexical items. According to these authors, HSs of Norwegian perform more accurately with grammatical gender on declensional suffixes (e.g., *hest<u>en</u>*, 'the horse'), which may be stored alongside nominal roots as single units in the mental lexicon, than on independent, indefinite articles (e.g., *<u>en</u> hest*, 'a horse'), which "must be

computed as part of a productive process" (p. 11). The intuition here seems to be the following. Abstract features, such as grammatical gender, are produced less variably when they can be retrieved from the lexicon as a single unit. When abstract features are instantiated on separate lexical items, as in the case of the indefinite articles, then HSs must perform a process of syntactic agreement, which is somehow more "complex" than simply retrieving a single unit from the lexicon. (For a similar example involving discontinuous and continuous negation in Arabic, see Albirini & Benmamoun, 2015).

A second potential line of inquiry into HSs' lexical knowledge involves priming (see Pickering & Ferreira, 2008 for a review of structural priming research). If HSs' productive variability is attributable, at least in part, to difficulty in accessing functional features from the HL lexicon, then perhaps exposing them to these features immediately prior to production could decrease the variability that they demonstrate. Dorian (1981) reports that she was unable to elicit a certain complex construction from a Gaelic HS until she provided that HS with a sample sentence (presumably activating dormant featural knowledge), at which point she began to produce this structure repeatedly and in a target-like manner. This evidence notwithstanding, there is no research to my knowledge which explores priming as a predictive factor of HS variability.

2.3.4. Summary of approaches to HS divergence and variability

In this chapter, I have outlined three broad approaches to understanding HS divergence and variability: The Input Quality approach, the Representational Differences approach, and the Activation/Lexicalist approach. Though I presented each of these approaches separately, it is important to clarify that they are not mutually exclusive.

Proponents of the Input Quality approach do not attribute *all* HS variability to input quality (e.g., Pascual y Cabo & Gómez Soler, 2015) and do not rule out the possibility of underlying representational differences. Similarly, proponents of the Representational Differences approach carefully consider both input quality (e.g., Montrul & Sánchez Walker, 2013) and lexical knowledge (Montrul et al, 2014) as potentially predictive factors in HS variability. Finally, proponents of the Activation/Lexicalist approach (Putnam & Sánchez, 2013) accept the possibility that HSs' variability can be caused by representational differences and/or input quality.

These similarities aside, I believe that the Activation/Lexicalist approach most clearly stimulates further research into the persistent variability demonstrated by HSs, opening up the chance to explore the potentially predictive roles of lexical storage, lexical frequency and featural priming in determining HSs' variability and divergence.

In Chapter 3, I will introduce subjunctive mood, the property of the interest in the present study, and some details about its syntax and acquisition by native speakers.

CHAPTER 3:
SUBJUNCTIVE MOOD: SYNTAX AND ACQUISITION

3.1. Introduction

In this dissertation project, I will be exploring HS divergence and variability, as presented in Chapter 2, by investigating HSs' productive and receptive knowledge of two types of subjunctive mood morphology in Spanish. In the present chapter, therefore, I will introduce the topic of subjunctive mood morphology in detail, setting the stage for the remainder of the dissertation.

The rest of Chapter 3 will be organized as follows. In Section 3.2., I will illustrate semantic and syntactic differences between two types of subjunctive mood: (a) intensional subjunctive (3.2.1), which is triggered by preceding lexical items, and (b) polarity subjunctive (3.2.2), which is triggered by preceding context. In Section 3.3., I will review previous research on the acquisition and maintenance of subjunctive mood by native speakers of Spanish, highlighting both (a) patterns of HSs' mood divergence and variability and (b) the effect of subjunctive type on HS knowledge.

3.2. Types of subjunctive mood

The term subjunctive mood refers to a broad class of morphological inflections in Spanish used to mark modality (Bosque, 2012). What all subjunctive mood inflections have in common is their formation, which always takes one of two possible shapes.

In most cases, subjunctive mood is formed exclusively by a shift in thematic vowel, as illustrated by the 3rd person singular verb forms in the first row of Table 1. The indicative mood forms of the verbs *comer* ('eat') and *hablar* ('speak'), *com-e* and *habl-a*, end with the

thematic vowels *–e* and *–a,* which mark their verb class. Their subjunctive mood counterparts, *com-a* and *habl-e*, share the same verb roots, *com-* ('eat') and *habl-* ('speak'), but differ in their vocalic endings, which shift from either *–e* to *–a* in the case of "–ar verbs" or or *–a* to *–e* with "-er/ir verbs" (Pérez Saldanya, 2012).

In other cases, subjunctive mood is marked by both a shift in thematic vowel and a change in the verbal root, as illustrated by the 3rd person singular verb forms in the second row of Table 1. The indicative forms of the verbs *tener* ('have') and *decir* ('say'), *tien-e* and *dic-e*, end with the thematic vowel *–e*, which marks them as so-called "er/ir verbs." Their subjunctive mood counterparts, however, are different in terms of both (a) vocalic endings, which shift from *–e* to *–a*, and (b) verbal roots, which shift from *tien-* to *teng-* in the case of *tener* and *dic-* to *dig-* in the case of *decir*. Verbs like *tener* and *decir*, therefore, seem to be doubly marked for subjunctive.

Table 1.
Mood morphology in Spanish

	Indicative Mood	Subjunctive Mood
I. Shift in Thematic Vowel	a. *com-e* ('eat') b. *habl-a* ('speak')	a. *com-a* ('eat') b. *habl-e* ('speak')
II. Shift in Thematic Vowel Plus Root Change	a. *tien-e* ('have') b. *dic-e* ('say')	a. *teng-a* ('have') b. *dig-a* ('say')

The fact that there is a unified set of morphological inflections that we derive in these two ways and call "subjunctive", however, obscures the multiple syntactic and semantic contexts that give rise to this class of verbal morphology. As Quer (2006) points out in his review of subjunctive syntax, "subjunctive does not constitute a syntactically uniform object, either crosslinguistically or even within the same language" (p. 661).

The present study focuses on two different structures that give rise to subjunctive mood: intensional subjunctive and polarity subjunctive. The remainder of Section 3.2.,

therefore, presents an overview of each of these subjunctive types, focusing on the meanings of these forms as well as the ways that those meanings are represented syntactically in the minds of Spanish speakers.

3.2.1. Intensional Subjunctive Mood

In intensional subjunctive structures, subjunctive mood morphology is triggered by a preceding lexical item or items, typically, though not necessarily, strong intensional predicates (Quer, 2001). In (1), it is the strong intensional predicate *quiere* ('want') that triggers the presence of subjunctive mood in the embedded clause. In (2), by contrast, the weak intensional predicate *creer* ('believe') in the main clause does *not* trigger subjunctive mood in the embedded clause.

(1) John <u>quiere</u> que Myra prepare/*prepara una torta para sus nietos.
John want [3psg] that Myra prepare [3psg-SUBJ/*IND] a cake for their grandchildren.
"John wants Myra to prepare a cake for their grandchildren."

(2) John <u>cree</u> que Myra *prepare/prepara una torta para sus nietos.
John believes[3psg] that Myra prepare [3psg-*SUBJ/IND]a cake for their grandchildren.
"John believes Myra is preparing a cake for their grandchildren."

In a sense, the subjunctive morphology on the verb *prepare* in (1) contributes no meaning to the sentence. After all, if a speaker were to use $prepara_{IND}$, instead of $prepare_{SUBJ}$, the resulting ungrammatical sentence would not have a different meaning from the grammatical version of (1). In another sense, however, the subjunctive mood morphology in (1) *does* contribute meaning, signaling a shift in the speaker's worldview. Quer (2001) argues that subjunctive mood morphology, triggered by strong intensional predicates like *quiere*, indicates a shift from the epistemic model of the speaker ($M_E(x)$) to

a model of buletic alternatives ($M_{BUL}(x)$). Whereas the epistemic model "stands for the speaker's worldview" and "represents…what s/he knows and believes," the buletic model introduces "a set of worlds that model alternative realizations of the actual world according to preferences of the matrix anchor" (p. 85). Under this account, the subjunctive mood in the embedded clause (proposition) of (1) can be seen as pointing to a shift from an *epistemic* perspective (e.g., John sees that Myra is actually preparing cake) to a *buletic* perspective (e.g., John imagines a world in which Myra is preparing a cake).

While most intensional subjunctive mood is triggered by strong intensional predicates, such as *querer*, it is also possible for other types of lexical items to trigger intensional subjunctive, including nouns, adjectives and prepositions (Bosque, 2012). Most relevant for the present study is the case of preposition, *para* ('for'), which together with the complementizer *que* ('that') obligatorily triggers subjunctive mood morphology in purpose clauses such as (3). Like *quiere* in (1), the lexical item *para que* triggers subjunctive mood in its embedded clause, which, in turn, signals a shift from an epistemic model to a model of buletic alternatives.

(3) Gramps$_i$ saca la basura <u>para que</u> Mame le$_i$ compre/*compra helado
Gramps take out[3psg] the trash for that Mame CL-3psg buy-3psg-SUBJ/*IND ice cream.
"Gramps takes out the trash so that Mame buys him ice cream."

Just as not all intensional predicates trigger subjunctive (e.g., *creer* in (2)), though, it is also true that not all complementizers[13] trigger subjunctive, as illustrated in (4), where the complementizer *porque* ('because') is followed by indicative mood.

[13] For the purposes of this explanation, *para que* is analyzed a single complementizer. No crucial theoretical implications hinge on this decision.

(4) Gramps$_i$ saca la basura porque Mame le$_i$ *compre/compra helado
Gramps take out [3psg] the trash because Mame CL-3psg buy-3psg-*SUBJ/IND ice cream.
"Gramps takes out the trash because Mame buys him ice cream."

Before exploring the syntactic representation of intensional subjunctive, it is important to first address a lingering question, namely, why indicative mood appears in the embedded clauses of sentences (2) and (4). There are two strong possibilities, each of which I will briefly consider here. The first possibility is that indicative mood is *triggered* by preceding lexical items, as in the case of the subjunctive mood morphology in sentences (1) and (3). It may be the case, for example, that the lexical item *porque* automatically triggers indicative mood morphology. The second possibility is that indicative mood is the default mood morphology in Spanish, meaning that it appears whenever there is an absence of specific lexical items that trigger subjunctive mood. In other words, if there is *not* a subjunctive trigger word like *para que*, then indicative mood appears.

Citing the observation that only indicative mood forms can appear in both embedded and main clauses[14], Bosque (2012) and Quer (2001) argue that indicative mood is the default mood morphology in Spanish, a hypothesis that will have considerable ramifications for both syntactic analyses of intensional subjunctive mood and, ultimately, the explanation of the results in the present study. If intensional subjunctive mood is triggered by a subset of lexical items and intensional indicative mood emerges in the

[14] Subjunctive mood can, however, appear in main clauses after certain epistemic adverbials such as *quizás* ('perhaps') and *probablemente* ('probably'), as noted by García (2011) and references therein. It is important to note, however, that subjunctive mood cannot appear in main clauses without epistemic adverbials, as shown in (i) and (ii):

(i) *quizás yo venga* ('perhaps I will come$_{SUBJ}$')
(ii) **yo venga* ('I come$_{SUBJ}$')

absence of these lexical items, then it may be best to reconceptualize intensional subjunctive morphology as little more than a reflection of the semantic content of the lexical items that trigger it, rather than a primary semantic contributor. In sentence (1), therefore, it is not the subjunctive morphology, *per se*, that signals a shift in model base (from epistemic to buletic) as much as it is the trigger verb *quiere*, whose lexical semantics require, and are reflected in, the mood morphology of the embedded verb *prepare*.

Kempchinsky (2009) proposes a syntactic analysis of intensional subjunctive that fits neatly with this conceptualization of subjunctive mood morphology as a *reflector* of preceding lexico-semantic information rather than a *contributor* of new semantic information. To translate Quer's (2001) observations about model shift (e.g., epistemic to buletic models) into the syntax, Kempchinsky proposes a new syntactic feature, W (worlds), which signifies "one particular world or a set of worlds" and "may come in both interpretable and uninterpretable guises" (p. 1798). In the case of intensional subjunctive, it is the uninterpretable instantiation *u*W that takes center stage.

In Kempchinsky's analysis of intensional subjunctive, represented below in (5), a lexical selector with a feature W selects for a CP with a ForceP headed by the uninterpretable feature, *u*W, a process Kempchinsky calls selection or identification. In Kempchinsky's proposed structure, the lexical selector is represented as a strong intensional predicate V such as *querer*. Nonetheless, it is quite possible to imagine other types of lexical items (e.g., prepositions) playing the same role as the selecting verb in (5). For intensional subjunctive with *para que*, for example, the lexical selector could be the preposition *para*, which would select for the same type of CP as in (5).

(5) …V_W [$_{CP}$[$_{ForceP}$ Force$_{[uW]}$] [$_{FinP}$ [$_{Fin}$ Op] [$_{IP}$ (DP) [$_{MoodP}$ [V+T+M_W] [$_{TP\ldots}$]]]]

Regardless of the selector's lexical class, the uninterpretable *u*W feature in Force would then be checked and deleted, via Agree, by the W feature in the complex Mood head.

Under this account, the default status of indicative mood is conceptualized by means of the W feature, whose default value is set at "W_R, the actual world according to the speaker" (p. 1799). When there is no triggering lexical item, such as *querer* or *para que*, Kempchinsky argues that the W_R feature "need not be identified and therefore will not necessarily be present in the syntax" (p. 1799). The reason that sentences (2) and (4), above, do not have subjunctive, therefore, is due to the absence of triggering, W-marked lexical items, rather than any indicative-triggering feature.

Before concluding Section 3.2.1, it is important to first comment on the question of dialectal variation in the use of intensional subjunctive mood in Spanish. The 81 participants in the present study, as we will see in Chapter 4, come from a wide variety of dialectal regions within the Spanish-speaking world. Consequently, major differences in the use of subjunctive mood across dialectal regions, especially with the complementizer *para que*, could potentially have a significant influence on the analysis and interpretation of data in the present study.

Gudmestad (2010) points out that relatively little is known about how subjunctive mood selection and use varies across different dialects of Spanish. Nonetheless, there are a few studies which have examined this topic in recent years. Blake (1982) found evidence of variability in Mexicans' and Spaniards' written use of mood morphology after lexical triggers such as *dudar* ('doubt') which, like *querer*, is expected to obligatorily select for

subjunctive. Spaniards in the study produced subjunctive 94% of the time with *dudar* while Mexicans used subjunctive only 72% of the time with *dudar*. Blake's study did not, however, test subjunctive use with *para que*.

Gallego & Alfonso-Marks (2014), in a comparative study of Spaniards in Toledo and Argentineans in Rosario, found evidence of slight dialectal variation in speakers' acceptance and preference of subjunctive and indicative mood after the lexical selector *querer* ('want'). While neither group categorically rejected indicative mood after *querer*, the Argentineans showed a markedly higher acceptance of this traditionally "ungrammatical" structure. Once again, however, the authors did not evaluate mood selection with the complementizer *para que.*

To my knowledge, only two studies have evaluated potential variation in monolinguals' use of subjunctive mood morphology with *para que*. Kowal (2007), in a study of mood variation in Costa Rica, reports a single instance of a native speaker using indicative mood after *para que*. Because the study was relatively small and included only four total instances of *para que*, it is impossible to say whether this unexpected observation is a sign of larger intensional subjunctive variation or a simple "performance error." Lastra & Butragueño (2012), on the other hand, point out, in accordance with the literature, that for monolingual Spanish speakers in Mexico City, certain adverbial triggers, such as *para que*, do not seem to *ever* select for indicative mood morphology.

More important, perhaps, than mood variation in *monolingual* contexts is the possible presence of mood variation in *bilingual* contexts, given that it is in these settings where HSs are exposed to their primary Spanish input. In a large-scale corpus study, conducted in New York City, Martillo Viner (2017) found that 1st generation immigrants

(age of arrival: 16 years or later) from Puerto Rico, the Dominican Republic, Cuba, Mexico, Ecuador and Colombia[15] did not ever produce indicative mood in 134 purpose clauses with *para que*. Based on this finding, it seems reasonable to conclude that *para que* obligatorily selects for subjunctive mood morphology across a variety of Spanish dialects, including dialects spoken by first-generation Spanish speakers living in the United States.

3.2.2. Polarity subjunctive mood

In the previous section, we examined intensional subjunctive mood, which is (a) triggered by lexical selection and, consequently, (b) reflects, rather than contributes, meaning to the sentence. In this section, we will explore polarity subjunctive mood, which differs from intensional subjunctive mood in both of these key respects.

(6) Kim$_i$ busca una raqueta de tenis que mejore/mejora su$_i$ saque.
Kim look for [3psg] a racquet of tennis that improve-3psg-SUBJ/INDIC her serve.
"Kim is looking for a tennis racquet that improves her serve"

Sentence (6) illustrates polarity mood selection and some of the ways in which it differs from intensional mood selection. First of all, note that the mood morphology on the verb *mejorar* ('improve') in the relative clause is not triggered by preceding lexical items. In his discussion of mood morphology in restrictive relative clauses such as (6), Quer (2001) notes that "the issue of lexical selection does not arise, because relative clauses are not selected by the main predicate" (p. 90). Unlike the strong intensional predicate *quiere* in (1), which selects for one particular type of CP (e.g., a CP with a [*u*W] feature in ForceP),

[15] As we will see in Chapter 4, 24 of the 42 HSs in this dissertation project (57.1%) grew up with parents from at least one of these countries, making Martillo Viner's (2017) finding particularly relevant here.

the matrix verb *busca* ('looks for') in (6) does not impose any selectional restrictions on the structure of the relative clause that modifies its object. Second, note that both subjunctive and indicative mood morphology are grammatically possible in the relative clause in (6), albeit with different meanings. In this respect, too, polarity subjunctive differs from intensional subjunctive, where lexical selectors such as *quiere* and *para que* can only be followed by subjunctive (and never indicative) mood morphology.

How does mood morphology in the relative clause impact the meaning of (6)? When the indicative mood form, *mejora*, is used, the DP (in this case, the tennis racket) is interpreted as specific or "referential." In (6), for example, this might mean that the subject *Kim* is looking for a particular tennis racket, which she knows to exist (e.g., it is her racket, and she cannot find it; she borrowed it once from a friend and noted how it helped her serve…etc…), that improves her serve. When the subjunctive mood form, *mejore*, is used, however, the DP is interpreted as "non-referential" (Borgonovo, Bruhn de Garavito & Prévost, 2015) or "attributive." In (6), this might mean that *Kim* is looking for a racket that will improve her serve, but she does not know if such a racket exists. The difference in mood forms, therefore, comes down to presupposition. When the speaker presupposes the existence of the DP, indicative is used in the relative clause. In the absence of presupposition, however, subjunctive mood is used.

Quer (2001) accounts for this mood alternation by making reference, once again, to shifts in a speaker's model evaluation. From this perspective, indicative mood is used in relative clauses when "the property in question must be evaluated…in the epistemic model of the speaker M_E(speaker)" while subjunctive mood is used to signal that "the property expressed by the relative has to be evaluated in M_{BUL}(they)" (p. 90). For Quer, polarity and

intensional subjunctive morphology, though different, are both used to mark shifts in a speaker's models of evaluation.

If both indicative and subjunctive mood forms are grammatical in adjectival relative clauses such as (6), how can we determine whether a speaker's use of mood is target-like? The key, of course, is context, which can be shaped by different factors, including (but not limited to) the lexical semantics and grammatical aspect of the matrix verb. Sentence (7) below is a slight alteration of (6). Whereas the matrix verb in (6) is the intensional verb *buscar*, the matrix verb in (7) is the non-intensional verb *encontrar* ('find'), whose meaning alone suggests presupposition. (It is not possible to find something whose existence you do not presuppose.) In addition, the verb is inflected with perfective morphology, suggesting that the action of finding, in this case finding the tennis racket, is complete. As a result of these two factors, the subjunctive form *mejore* in (7) becomes infelicitous.

(7) Kim$_i$ encontró una raqueta de tenis que #mejore/mejora su$_i$ saque.
Kim find [3psg-PERF] a racquet of tennis that improve-3psg-#SUBJ/INDIC her serve.
"Kim found a tennis racquet that improves her serve"

Similarly, it is possible to modify *contextual* factors in such a way that makes indicative mood forms infelicitous in such adjectival relative clauses. Imagine, for example, that Kim, the subject of (6)-(7), is in a large tennis equipment store, where she has been practicing her serve with a number of different tennis rackets. After an hour of trying out rackets, she cannot find a single one that improves her very slow serve. (Unfortunately, she is also not aware of any racket that is made to help tennis players improve their serves.) As Kim walks frustratedly through the aisles, a store clerk

approaches her and asks what she is looking for. In response to the clerk's question, Kim responds with (8), a slight alteration of (6) above.

(8) Busco una raqueta de tenis que mejore/#mejora mi saque.
Look for [1psg] a racquet of tennis that improve-3psg-SUBJ/#INDIC mi serve.
"I am looking for a tennis racquet that improves my serve"

In this invented context, where Kim clearly does not presuppose the existence of the tennis racket in question, indicative mood morphology in the relative clause becomes infelicitous or perhaps even ungrammatical. Clearly then, the acceptability of indicative and subjunctive mood morphology in adjectival relative clauses is dependent on the presence or absence of presupposition, which can be shaped by lexical semantics, grammatical aspect or context itself.

How is polarity subjunctive mood represented in the syntax? Recall that Kempchinsky (2009) posited the existence of a W (worlds) feature which can be either uninterpretable, as in the case of intensional mood structures, or interpretable. It is the interpretable instantiation of this feature, however, which figures prominently in Kempchinsky's analysis of polarity subjunctive syntax, represented below in (9). Note that two relevant differences between this structure and the proposed intensional subjunctive structure (presented above in (5), reprinted as (10)) are bolded and underlined.

(9) …**<u>V</u>** $[_{CP}[_{ForceP}\ \text{Force}_{\underline{\mathbf{[W]}}}]\ [_{FinP}\ [_{Fin}\ \text{+Fin}]\ [_{IP}\ (\text{DP})\ [_{MoodP}\ [\text{V+T+M}_{W}]\ [_{TP\ldots}]]]]]$ (Polarity)

(10) …V_{W} $[_{CP}[_{ForceP}\ \text{Force}_{[uW]}]\ [_{FinP}\ [_{Fin}\ \text{Op}]\ [_{IP}\ (\text{DP})\ [_{MoodP}\ [\text{V+T+M}_{W}]\ [_{TP\ldots}]]]]]$ (Intensional)

Before entering into the syntactic details, it is important to highlight that (9) represents Kempchinsky's attempt to model polarity subjunctive in negated epistemics (e.g., *No creo que mis estudiantes hagan la tarea*, 'I don't believe that my students do the homework'), rather than in restrictive relative clauses such as (6) to (8). Nonetheless, it is easy to see how this analysis could be extended to relative clauses by simply removing the initial verb, V, and assuming that the resulting CP is an adjectival relative clause.

Unlike intensional subjunctive morphology, which simply *reflects* the lexical semantics of the subjunctive trigger that precedes it, polarity subjunctive morphology contributes meaning to the sentence, a fact that Kempchinsky builds into her proposed polarity subjunctive structure. Because there is no lexical selector in polarity subjunctive structures, note that the verb, V, at the beginning of Kempchinsky's proposed polarity structure in (9) is no longer marked with the W feature. Instead, Kempchinsky proposes that the interpretable W feature emerges in the Force head. Since this feature is interpretable, unlike the *u*W feature in the Force head of intensional subjunctive structures, it is not checked. On the contrary, "the relationship between W in Force and the verbal complex is…identification" (p. 1799).

To conclude Section 3.2.2., it is important to briefly consider differences in the process of producing, as opposed to comprehending, polarity subjunctive mood forms.

In production, Spanish speakers use mood morphology in relative clauses to signal a lack of presupposition. In isolated, out of the blue contexts, such as (6) above, the use of subjunctive or indicative mood is a significant tool that speakers can use to express this difference. In other cases, such as (7) or (8), speakers' use of subjunctive or indicative mood in the relative clause is just one of a few pieces of information pointing towards a presence

or absence of presupposition. In (7), the lexical semantics of the verb *encontrar*, coupled with its perfective grammatical aspect, clearly signal that the speaker presupposes the existence of the tennis racket in question, regardless of the mood morphology in the relative clause. In (8), contextual indicators (e.g., the speaker's inability to find the racket) point towards a lack of presupposition even *before* the speaker uses subjunctive morphology in the relative clause. It is not the case, therefore, that mood morphology is the only way in which Spanish speakers can conceivably express the presuppositional status of a given DP.

In comprehension, Spanish speakers decode other speakers' messages, rather than generating their own. In out of the blue contexts, such as (6), the listener must rely on mood morphology to intuit whether the speaker is expressing a presence or absence of presupposition. In cases like (7) and (8), however, where lexical semantics, grammatical aspect and other contextual indicators enter the equation, it may be possible for listeners to *comprehend* the speaker's expression of (non)presupposition without necessarily recognizing the mood morphology on the verb in the relative clause.

To conclude Section 3.2.2., I will now comment briefly on the possibility of dialectal variation in the use of polarity subjunctive mood in adjectival relative clauses. To my knowledge, there are no studies which have tested dialectal variation in monolingual Spanish speakers' use of subjunctive mood in adjectival relative clauses. There is one study, however, which has explored adjectival relative clause subjunctive (as well as a number of other subjunctive mood structures) in first-generation immigrants living in the US.

Bookhamer (2013), in the same corpus used in Martillo Viner's (2017) study (see Section 3.2.1), found no significant effects of dialect region on first-generation immigrants' use of polarity subjunctive mood in adjectival relative clauses. Nonetheless, he did find

that first-generation immigrants only used subjunctive mood 84% of the time (134 of 160) in non-presuppositional adjectival relative clauses with *que*, a finding that appears to contradict descriptions of these clauses in the literature. It is important to note, though, that Bookhamer's data consists of spontaneous interviews, rather than controlled experiments, which are analyzed and coded by the researcher. With such open-ended data, it may be harder for the researcher to reliably identify whether the speakers are really presupposing (or not) the existence of the NPs that they modify with adjectival relative clauses.

3.3. Acquisition of subjunctive in Spanish as a native language

Thus far, I have briefly reviewed the structure and meaning of intensional and polarity subjunctive mood forms in Spanish. In this section, I will review previous research on the acquisition of subjunctive by three types of native Spanish speakers (including HSs; see Rothman & Treffers-Daller, 2014 for discussion): Spanish monolingual L1 children (3.3.1), Spanish-English bilingual (2L1) children (3.3.2), and, finally, adult HSs (3.3.3).

In each of these three subsections, I will achieve the following goals. First, I will discuss the learning task faced by each of these groups in the acquisition and/or maintenance of subjunctive mood, based on the analyses of subjunctive mood presented in Section 3.2. Second, I will illustrate differences in these groups' exhibited knowledge of intensional, as opposed to polarity, subjunctive mood. Finally, I will highlight evidence of both divergence and variability in these groups' subjunctive mood knowledge, focusing on how the three approaches to HS divergence/variability presented in Chapter 2 are able to account (or not) for the patterns observed herein.

3.3.1. Acquisition of subjunctive mood: L1 children

In order to best understand adult HSs' knowledge of subjunctive mood, it is critical to first understand monolingual children's developmental trajectory with this property. After all, many adult HSs of Spanish, including quite a few in the present dissertation project, are *sequential HSs*, meaning that they were functionally monolingual Spanish speakers until beginning school at age 5 or 6. Tracing monolingual children's subjunctive development, therefore, provides insight into the linguistic knowledge that sequential HSs (likely) had prior to undergoing the dominance shift that characterizes HL acquisition (e.g., Austin, Blume & Sánchez, 2015).

3.3.1.1. The L1 learning task

What is the learning task for monolingual children learning subjunctive mood in Spanish? With intensional subjunctive mood, they must (i) acquire the uninterpretable *u*W feature in ForceP (Section 3.2.1) and (ii) learn the W-marked lexical items (e.g., *querer*, *para que*…etc…) that select for a ForceP with that feature. Both of these steps are critical elements of the intensional subjunctive learning task. In Kempchinsky's (2009) analysis, the uninterpretable *u*W feature is the prime mover in the production of intensional subjunctive, given that this feature's uninterpretable status requires that it be checked and deleted via Agree with the W feature of the Mood head. Without acquiring this abstract feature, children would not produce intensional subjunctive forms because there would be no underlying syntactic motivation for doing so.

Acquiring the *u*W feature, however, by no means *guarantees* target-like knowledge of intensional subjunctive mood, given that children must also learn the specific lexical

items that select for complements with this feature in the head of ForceP. In the process of learning intensional subjunctive, therefore, it may be the case that children pass through a stage in which they have acquired the *u*W feature but only learned some of the lexical items that select for it. As children learn which lexical items select for a ForceP marked by the *u*W feature, they may exhibit asymmetric intensional subjunctive patterns, e.g., producing subjunctive after the lexical selector *querer* but not after the lexical selector *dudar* ('doubt'). Ultimately, they may come to store different lexical selectors such as *para que* alongside their complements as treelets (Jackendoff, 2007) in the mental lexicon, facilitating their access to, and retrieval of, the *u*W feature for production and comprehension purposes.

Even after children have acquired the *u*W feature *and* learned the W-marked lexical items that select for it, they *still* may exhibit non-target production and/or recognition of subjunctive mood since knowing subjunctive also means learning the ways in which its abstract features are instantiated morphologically on different lexical items. Imagine, for example, a child who has acquired the relevant abstract features *and* learned all of the triggering lexical selectors. In a given day, let's say that this child produces sentences (11)-(13), all of which start with the W-marked lexical selector *quiero* but only two of which are marked with target subjunctive morphology.

(11) Quiero que compres helado
Want [1psg] that buy-SUBJ-[2psg] ice cream.
"I want you to buy ice cream."

(12) Quiero que corras rápido
Want [1psg] that run-SUBJ-[2psg] fast.
"I want you to run fast."

(13) *Quiero que lees el libro.
Want [1psg] that read-*INDIC-[2psg] the book.
"I want you to read the book."

In (11) and (12), the child produces target subjunctive mood morphology, seemingly showing that he has acquired the *u*W feature and learned that *quiero* is one of its many triggers. In (13), however, the child produces the non-target indicative mood form *lees* ('you read') instead of the expected subjunctive form *leas*. What could be the cause of such variability in a child's knowledge of intensional subjunctive mood? One possibility is that the child has simply not learned that the second person, subjunctive marked form of *leer* is *leas*. That is to say, the child exhibits variability because he is in the process of learning the ways in which abstract subjunctive features are mapped onto individual verbal stems. If this were the case, it is almost inevitable that the child would exhibit some intra-speaker variability in development as he gradually expands his knowledge of what subjunctive mood looks like on different verb forms.

The learning task for polarity subjunctive is similar in many ways, though certainly more complex. First, children must learn the interpretable W feature in ForceP, which, according to Kempchinsky's (2009) analysis, undergoes an identification relationship with the W feature of the Mood head. This time, however, the child must learn to associate the feature not to a lexical item, such as *para que*, but to a contextual status, e.g., the lack of presupposition (or in Quer's terminology, a shift in model evaluation). This difference makes the task of learning polarity subjunctive more difficult for two reasons.

The first reason learning polarity subjunctive is more difficult has to do with lexical storage. It is likely, as argued above, that lexical triggers for intensional subjunctive such as *para que* are stored as treelets (alongside their complementary structure) in the mental lexicon of Spanish speakers. Therefore, when a speaker activates the lexical item *para que*, it likely results in the automatic, co-activation of its complementary structure, including

the *u*W feature in the head of ForceP. However, because the W feature driving polarity subjunctive is associated with a presuppositional status, rather than a specific triggering lexical item, it is harder to imagine how this feature might be stored as a facilitative treelet in the minds of Spanish speakers. From the standpoint of lexical storage/access, therefore, intensional subjunctive forms are likely easier to retrieve and produce, given that their relevant abstract features are mapped reliably and predictably to specific lexical selectors.

The second reason why the polarity subjunctive learning task is more difficult, at least for children, is because its acquisition requires speakers to have reached a certain level of cognitive maturity. In order to recognize presupposition (or its absence), which drives mood selection in relative clauses, children may need to be cognitively mature enough to recognize other perspectives (Pérez-Leroux, 1998), as we will discuss in the coming paragraphs. This factor alone makes the learning task for polarity subjunctive, at least in relative clauses, far more difficult than that of intensional subjunctive.

Despite these crucial differences, the last element of the polarity subjunctive learning task is identical to that of the intensional subjunctive learning task. Even after the interpretable W feature has been acquired and speakers have learned to associate it with the absence of presupposition in relative clauses, they must still learn the ways in which subjunctive mood is instantiated on specific lexical items. As argued above, and demonstrated in (11) to (13), it is conceivable that children learning polarity subjunctive mood may produce it more with some lexical items than others as their knowledge of subjunctive mood instantiations grows over time.

3.3.1.2. Findings from L1 acquisition research

A survey of the L1 acquisition literature indicates that intensional subjunctive with *para que* is acquired before polarity subjunctive mood in adjectival relative clauses.

López Ornat (1994) reports an example of an L1 child producing subjunctive with *para que* as early as 2;06, as presented in Montrul (2004a) and shown in (14).

(14) Esto es para ti, para que te lo eches (María: 2;06)
This is for you, so that CL-refl. it throw-SUBJ-[2psg]
"This is for you to put on."

(López Ornat, 1994)

The relatively early acquisition of intensional subjunctive forms is also corroborated by the research of Blake (1983), who studied the subjunctive knowledge of monolingual children between the ages of 4 and 12. Of the subjunctive structures tested in Blake's study, children performed most accurately with two intensional subjunctive forms: volitional subjunctive with *querer* and purpose-clause subjunctive with *para que*.

Not surprisingly, children appear to have more difficulties with polarity subjunctive in relative clauses than they have with intensional subjunctive in purpose clauses. Pérez Leroux (1993) tested monolingual children's knowledge of both intensional and polarity subjunctive mood forms using an innovative elicited production task called *Busca, Busca* ('Look, look'). Participants in the study, who ranged in age from 3;2 to 6;10, were presented with short scenarios, such as (15) below, in which a character is searching for something. At the end of each scenario, the experimenter asked the participants to describe what the character is looking for. Because the characters in each scenario are unable to find the

objects of interest (e.g., a knife that cuts meat), these contexts bias participants towards "an interpretation of the NP without presupposition of existence." (p.170). Consequently, participants with adult-like knowledge of polarity subjunctive are expected to produce adjectival relative clauses with subjunctive mood morphology.

(15) Prompt: *La cocinera tiene que cortar una carne. Coge un cuchillo pero no corta bien. Coge otro pero ese es para untar mantequilla.*
(The chef has to cut a piece of meat. She grabs one knife but it doesn't cut very well. She grabs another knife but that one is for spreading butter.)

Question: *¿Qué busca la cocinera?*
(What is the chef looking for?)

Expected Answer: *Un cuchillo que corte*$_{SUBJ}$ *la carne.*
(A knife that cuts$_{SUBJ}$ meat.)

(Pérez-Leroux, 1993, 1998)

The youngest subset of participants (3;2 to 3;11) barely produced any relative clauses at all, let alone relative clauses with subjunctive mood. However, the older participants demonstrated gradual development in their knowledge of polarity subjunctive. The five participants between the ages of 4;2 and 4;11 produced a total of 14 adjectival relative clauses: 6 (42.8%) with target-like subjunctive morphology and 8 (57.2%) with non-target indicative morphology. At this stage of development, it seems that monolingual children alternate evenly between indicative and subjunctive mood morphology in non-presuppositional adjectival relative clauses. The five participants between the ages of 5;2 and 5;11 performed in a more adult-like manner, producing target-like subjunctive mood morphology in 13 out of 18 (72.2%) adjectival relative clauses. Though the older children

generally performed more accurately, it is noteworthy that their production of subjunctive mood in non-presuppositional relative clauses was still not categorical.

Interestingly, the researcher also elicited intensional subjunctive forms from the same participants, making it possible to draw rough comparisons between their knowledge of intensional and polarity subjunctive mood. When participants in the experiment answered without specification (e.g., not modifying the object of interest: *un cuchillo* ('a knife') instead of *un cuchillo que corte la carne* ('a knife that cuts meat')), the experimenter sought to elicit more specific answers by asking follow-up questions with *para que* ('so that'), which triggers subjunctive, and *porque* ('because'), which does not.

In 49 follow-up questions with *para que*, participants answered with 100% accuracy, producing exclusively target subjunctive mood forms (e.g., *para que pinte*$_{SUBJ}$ *allí* ('so that she paints over there'); Veronica: 3;8)) or target infinitival forms (e.g., *para hacerle una pregunta* ('to ask him a question'); Gaby: 3;8). Participants in the youngest age range (3;2 to 3;11), despite barely producing any adjectival relative clauses at all, performed categorically with these intensional subjunctive forms. The participants' accuracy in these follow-up questions, however, was not limited to subjunctive or infinitival forms. In 19 follow-up questions with *porque*, participants again answered with 100% accuracy, producing 19 target-like indicative mood forms.

The fact that monolingual children, in the same experiment, produce intensional subjunctive forms categorically and polarity subjunctive mood forms variably is strong evidence that intensional subjunctive forms are acquired and mastered at an earlier age.

Pérez-Leroux (1998), in a similar study of polarity subjunctive mood using the same elicited production task, finds a strong relationship between cognitive maturity and

participants' ability to produce subjunctive in adjectival relative clauses. Specifically, children who passed a standard Theory of Mind (ToM) examination were over 19 times more likely to produce subjunctive relative clauses than children who did not pass the ToM examination. This strong correlation between cognitive maturity and polarity subjunctive production suggests that children's later acquisition of polarity subjunctive is at least partially attributable to difficulties grasping the semantic notion of presupposition.

In summary, previous L1 acquisition research reveals that monolingual children acquire intensional subjunctive mood forms quite early (e.g., before age 3) and polarity subjunctive mood forms much later (e.g., age 6 or later). In the next section, I will shift my focus to previous studies of the acquisition of subjunctive mood by early-childhood (simultaneous) bilinguals, whose acquisition of subjunctive takes a slightly different shape.

3.3.2. Acquisition of Subjunctive Mood: 2L1 Children

In the previous section, we reviewed research on the acquisition of subjunctive mood by monolingual Spanish-speaking children, whose knowledge is likely to approximate the subjunctive knowledge of sequential HSs just before starting elementary school in English. In this section, I will add to this overview by summarizing research on the acquisition of subjunctive by child bilinguals in the process of learning both English and Spanish during early childhood. As noted by Montrul (2004a), "very little is known about the development of subjunctive in early bilingualism" (p. 143). Nonetheless, the few studies that have examined this topic shed light on what simultaneous and early sequential HSs know about subjunctive mood during the earliest stages of language acquisition.

3.3.2.1. The 2L1 learning task

In principle, the intensional and polarity subjunctive learning tasks for 2L1 children should be quite similar to the learning tasks for L1 children. Like the monolinguals, bilingual children must acquire the abstract features *u*W and W, as well as the specific lexical items and contexts that modulate the presence of these features. Because bilingual children are also still maturing cognitively, it is expected that polarity subjunctive will be more difficult for them, just as it was for monolingual children. Finally, just like monolingual children, bilingual children must also learn how the abstract features that they acquire are instantiated on specific lexical items.

Despite these similarities, the learning task for bilingual children is different from the learning task for monolingual children in at least two key ways. First, bilingual children are almost certain to receive less overall input in Spanish, given that their linguistic experience is divided between two different languages (e.g., Hoff, 2006). Second, bilingual children have two different linguistic systems, raising the possibility that they will experience cross-linguistic influence (CLI) between their two languages (e.g., Muller & Hulk, 2001). In the paragraphs that follow, I will consider how these two differences could impact early bilinguals' acquisition of subjunctive mood morphology in Spanish.

How might reduced relative exposure to Spanish impact early bilinguals' acquisition of intensional and polarity subjunctive mood? The first possibility is that early bilinguals, due to reduced quantity of Spanish input, will simply not hear enough instances of intensional and/or polarity subjunctive mood in order to acquire stable knowledge of the abstract features *u*W and W. In principle, this possibility makes sense. As a hypothetical bilingual child's total Spanish input approaches "0," there must be a point at which total

exposure is so low that the child cannot possibly intuit abstract grammatical knowledge of subjunctive mood. Nonetheless, as pointed out by Putnam & Sánchez (2013), it is difficult to know how much exposure is necessary for a bilingual child to acquire a linguistic property, making this line of inquiry nearly impossible to follow.

A second, and far more measurable possibility is that early bilinguals' lower relative Spanish exposure will impact their *lexical* knowledge of subjunctive mood, e.g., how subjunctive mood is instantiated on different verbs. If early bilinguals hear less Spanish, then they will presumably receive less exposure to the subjunctive mood instantiations of certain lower frequency verbs. It is easy to imagine, for example, an early bilingual who acquires the *u*W feature but never learns its instantiation on a less frequent verb form, such as *observar* ('observe'), which he may have rarely, if ever, heard inflected with subjunctive mood morphology. The path towards categorical subjunctive production may be more difficult, therefore, for the early bilingual because he must learn the subjunctive instantiations of many verb forms despite having had less chances to hear them.

Bilingual children's task of learning subjunctive mood may also be impacted by the interaction of two qualitatively different languages within individual minds. Decades of research has provided robust evidence of cross-linguistic influence (CLI) in the minds of bilingual children (e.g, Muller & Hulk, 2001; Nicoladis, 2006; Sánchez, 2004; Yip & Matthews, 2000, 2007; *inter alia*). It is a distinct possibility, therefore, that CLI in the minds of Spanish-English bilingual children will significantly affect their task of acquiring intensional and polarity subjunctive mood forms, respectively. Sánchez (2004) reports that Quechua-Spanish child bilinguals experience "functional convergence," in which evidentiality features present in the Quechua Tense head become "associated with Spanish"

(p. 161), leading bilinguals to exhibit innovative uses of Spanish aspectual morphology. Sánchez's study reveals that bilinguals are likely to experience CLI when a particular functional head, in this case Tense, has different feature matrices in each of a bilingual's two languages. If the head of ForceP in English lacks the W and/or *u*W features found in Spanish ForceP's feature matrix, then English to Spanish CLI, of a type similar to that reported by Sánchez (2004), may make it more difficult for bilingual children to access the W and *u*W features which trigger the production of subjunctive mood in Spanish.

3.3.2.2. Findings from 2L1 Acquisition Research

As pointed out above, there is little research on early bilinguals' acquisition of subjunctive in Spanish. Nonetheless, there are three studies which touch on this topic, directly or indirectly, and therefore, are highly relevant to the present dissertation project.

Silva-Corvalán (2014), in a longitudinal study of two English-Spanish bilingual children growing up in Los Angeles, sheds light on how bilingual children may come to acquire intensional subjunctive forms with *para que*. Before age 3, both children appear to exhibit qualitatively divergent purpose clauses such as (16), produced by Brennan at 2;5.6. Notably, Brennan's sentence in (16) diverges structurally, rather than morphologically, from the adult form (17). At this stage at least, it appears to be the case that Brennan's *para que* selects for different complementary structure.

(16) Para el bicho no picar a yo
For the bug not sting to I
"So that the bug doesn't sting me"

(17) Para que el bicho no me pique
For that the bug no me sting-[3psg]-SUBJ
"So that the bug doesn't sting me"

As the boys get older, however, their production of intensional subjunctive mood in purpose clauses with *para que* takes a more adult-like shape, as illustrated in (18), produced by Nico at 2;11.4, and (19), produced by Brennan at 3;1. Together, sentences (18) and (19) indicate that children acquiring subjunctive must (independently) learn both (a) functional features and (b) the instantiations of those features on specific lexical items.

(18) para que el pasto crezque con la arena
so that the grass grow$_{SUBJ}$-IND with the sand.
"So that grass may grow with the sand."

(19) para [que] los niños puedan subir
so that the boys can-SUBJ-[3pp] climb
"So that the boys can climb."

In (18), Nico produces an adult-like intensional subjunctive structure with ambiguous subjunctive mood morphology. The verb *crecer* ('grow'), like many verbs in Spanish, is doubly marked for subjunctive mood, as its subjunctive mood forms exhibit both a root change (from *crec-* to *crezc-*) and a thematic vowel shift (from *–e* to *–a*). Nico's innovative verb *crezque*, however, exhibits only the root change and, crucially, not the thematic vowel shift. Why does such a form emerge in Nico's speech? One strong possibility is that Nico has acquired the *u*W feature, hence the production of the root *crezc-*, but not yet learned its full subjunctive instantiation. Sentence (18), therefore, is evidence that children mastering subjunctive mood morphology gradually learn the way that subjunctive mood is marked on different verbs in Spanish. Sentence (19), in which Brennan unambiguously marks subjunctive mood morphology on the verb *puedan*, shows that

bilingual children produce target-like intensional subjunctive mood forms as early as approximately three years of age.

While Silva Corvalán's (2014) longitudinal research illustrates the gradual emergence of target-like intensional subjunctive forms in bilingual children's speech, Merino's (1983) longitudinal research shows the instability of this subjunctive knowledge in the face of rapidly increasing English language input and use.

Merino tested Spanish-English bilingual children in California, ranging from kindergarten to fourth grade, on their knowledge of a number of grammatical properties in Spanish including gender, tense/aspect and mood morphology. To test participants' productive knowledge of these different forms, the researcher used a delayed imitation task in which children listened to a sentence and repeated it after a short pause. In the subjunctive mood morphology category, participants heard and then attempted to repeat two intensional subjunctive forms (with *para que* and *quiere que*, respectively) as well as a polarity subjunctive form (with *tal vez*, 'perhaps').

At the first testing session, participants were 70% accurate in the repetition of subjunctive mood. By the second testing session, however, participants' accuracy with subjunctive mood had dropped to 55%. Because Merino presents only a single, overall subjunctive accuracy score, it is not clear the extent to which this decreased accuracy affects intensional, as opposed to polarity subjunctive forms. Nonetheless, these results suggest that Spanish-English bilingual children likely begin to exhibit increased variability with intensional subjunctive forms during early elementary school (e.g., ages 5-7), perhaps as a function of increased usage of and exposure to English. Based on the results of a background questionnaire, which assessed participants' daily language use, Merino argues

that the bilingual children exhibiting the greatest overall increases in Spanish grammatical variability were those "who tended to use both languages with the majority of their relatives and friends" (p. 291). Intensional subjunctive variability, therefore, may be primarily driven by longitudinal decreases in bilingual children's Spanish use and input.

A third study to evaluate Spanish-English bilingual children's acquisition of subjunctive morphology is Anderson's (2001) study of the siblings Beatriz and Victoria, who moved to the mainland US from Puerto Rico at ages 3;6 and 1;6, respectively.

Data collection, which consisted of bi-monthly, 30-minute "Spanish-only" recordings, began when Beatriz was 6;7 and her younger sister Victoria was 4;7. Over time, both sisters began to experience increased variability (referred to as "errors" in Anderson's paper) with a variety of morphological inflections including, but not limited to, subjunctive mood. (Strikingly, Victoria did not produce a single instance of present subjunctive in the final five recording sessions.) Most of the bilingual children's subjunctive "errors" consisted of producing indicative mood morphology where subjunctive was expected. Unfortunately, because Anderson does not specify the types of syntactic structures in which this occurred, it is not possible to determine whether the children's increased subjunctive variability occurred more with intensional or polarity subjunctive forms, respectively. Nonetheless, this study clearly points to bilingual children's increasing variability with subjunctive mood morphology over time, at least in a minority language context.

Interestingly, Victoria's increased morphological variability co-occurred with a sharp decrease in lexical productivity. After producing an average of 27.8 different Spanish verbs/session over the first six recordings, Victoria produced just 19 verbs/session in the final six recordings. Victoria's reduced productive verbal repertoire suggests that over time,

she is experiencing increased difficulty accessing and producing certain verbs in Spanish. This pattern is consistent with the possibility presented in Section 3.3.2.1, namely that *some* of bilingual children's increased morphological variability over time may be attributable to either (a) losing or (b) never learning the exact instantiations of abstract features on certain, less frequent lexical items or verbal roots. Supporting this somewhat speculative claim is Anderson's observation that most of the children's morphological "errors" in the twelve recordings "were noted in the production of verbs that were low in occurrence and that were used only occasionally by the children" (p. 393).

3.3.3. Acquisition of subjunctive mood: adult HSs

In Sections 3.3.1 and 3.3.2, respectively, I reviewed previous research on monolingual and bilingual children's acquisition of subjunctive mood morphology in Spanish. The goal of these sections, broadly speaking, was to illustrate how children, in both monolingual and bilingual contexts, come to exhibit knowledge of subjunctive mood morphology in Spanish over time.

In the present section, I shift my focus to *adult* HSs of Spanish, who started out as either Spanish monolinguals (Section 3.3.1) or early Spanish-English bilinguals (Section 3.3.2) before experiencing a shift towards English dominance over time.

3.3.3.1. The adult HS learning and maintenance task

Because the present section focuses on *adult* HSs of Spanish, the subjunctive mood "learning task" presented in this section could just as easily (and perhaps more appropriately) be described as a "learning and maintenance task." Unlike monolingual and

bilingual children, whose subjunctive mood learning task consists of *acquiring* featural and lexical knowledge for the first time, adult HSs' "learning task" instead involves both (a) *preserving* previously acquired linguistic knowledge and, in some cases, (b) *expanding* that knowledge to newly learned lexical items.

To begin, let's imagine the case of an adult HS of Spanish who, during childhood, "successfully" acquired intensional subjunctive mood, learning both the uninterpretable *u*W feature and the lexical items (e.g., *para que*) that select for it. What would such a HS have to do maintain, and perhaps even expand, this knowledge of intensional subjunctive mood throughout adulthood?

It is likely, following Putnam & Sánchez (2013), that this HS would have to continue activating the HL for production purposes in order to maintain consistent access to his knowledge of the uninterpretable *u*W feature in ForceP. In addition to maintaining knowledge of and access to the *u*W feature, this HS would also have to maintain knowledge of the lexical items (e.g., *para que*) which select for complements with that feature. This maintenance task may be easier if the HS successfully stores the lexical selectors as treelets (Jackendoff, 2007) alongside their complementary structure in the mental lexicon.

Based on his exposure to and usage of the language, this HS may experience either an expansion or a reduction of his lexical knowledge of subjunctive mood morphology, leading to either decreases or increases in his exhibited subjunctive mood variability. Recall that Spanish speakers acquiring subjunctive must learn not only abstract features such as *u*W but also how those features are instantiated on certain lexical items. If this HS continues using Spanish in an increased range of contexts and registers, he may come to learn how subjunctive mood features are instantiated on an ever-larger set of Spanish

lexical items. If, on the other hand, this HS limits the range of contexts in which he uses Spanish, he may begin to lose knowledge of how subjunctive is instantiated on certain lexical items, resulting in increased mood variability.

The learning and maintenance task is similar, albeit slightly more complex, in the case of polarity subjunctive mood. Let's now imagine the case of an adult HS who "successfully" acquired polarity subjunctive during childhood, learning both the interpretable W feature and its association with a lack of presupposition in the context. What would such a HS have to do to maintain and expand this knowledge of polarity subjunctive mood throughout adulthood?

As in the case of intensional subjunctive, this HS would presumably have to activate the W feature (in relative clauses) enough to maintain access to it during speech production. Even if this feature is maintained, however, there is one key reason why maintaining knowledge of polarity subjunctive mood in relative clauses may prove more challenging for adult HS. Because the W feature is associated with a presuppositional status, rather than selecting lexical items, it may not be possible to store the W feature as a treelet in the mental lexicon alongside its complementary structure. (How could the W feature be stored in the lexicon with a "lack of presupposition," which is a contextual state rather than a lexical item?) This could have consequences for the ease with which adult HSs maintain the association between W and non-presupposition.

As in the case with intensional subjunctive, adult HSs would also have to maintain knowledge of how the W feature is instantiated on certain verb forms. As a result, increased or decrease use of Spanish lexical items over time could lead to lower or higher mood variability in polarity subjunctive structures.

3.3.3.2. Findings from adult HS research

Since Lantolf (1978) and Guitart (1982), researchers have examined HSs' knowledge of subjunctive mood from sociolinguistic (e.g., Bookhamer, 2013; Fernández Pedraza, 2014; Lynch, 1999; Martillo Viner, 2017; Martínez Mira, 2006; Ocampo, 1990; Silva Corvalán, 1991, 1994; *inter alia*), linguistic (e.g., Giancaspro, under revision; Montrul, 2007, 2009; Montrul & Perpiñán, 2011; van Osch & Sleeman, *forthcoming*; Pascual y Cabo, Lingwall, & Rothman, 2012; Perez Cortes, 2016), pedagogical (e.g., Hislope, 2001; Mikulski, 2006) and cognitive (Potowski, Jegerski & Morgan Short, 2009; Torres, 2013) perspectives, respectively.

In this section, I have chosen to focus on four representative linguistic studies which I feel are maximally relevant for the present dissertation project: (1) Montrul, 2009; (2) van Osch & Sleeman, *forthcoming*; (3) Perez Cortes, 2016; and (4) Giancaspro, under revision. For each of these four studies, I will (briefly) summarize relevant results, placing particular emphases on both (a) patterns of HS divergence and variability and (b) differences in intensional and polarity subjunctive mood. In addition, I will consider how the researcher(s) in each study account for the divergent/variable patterns that they observe using the three different approaches to HS divergence/variability presented in Chapter 2.

Montrul (2009) tested the productive and receptive subjunctive mood knowledge of 65 adult HSs in Illinois. Though the study included HSs of advanced, intermediate and low proficiency levels, I will focus on the results of the advanced and intermediate groups since these are the proficiency levels that I will be testing in this dissertation project.

Participants in the study completed three experimental tasks. In the oral elicitation task, participants were asked two broad questions which were "designed to elicit opinions in extended discourse and use of subjunctive forms" (p. 256). Results of the oral elicitation task revealed that advanced and intermediate proficiency HSs exhibited variability in the oral production of subjunctive mood, unlike the Spanish-dominant control group. Nonetheless, both the advanced HSs (92.3%) and the intermediate HSs (80.7%) were well above chance in the production of subjunctive mood in expected subjunctive contexts.

In the morphology recognition task, participants read a short letter from a patient to a doctor. In the letter, there were 30 underlined verb forms: 15 where indicative was expected and 15 where subjunctive was expected. Though all 15 expected subjunctive items were "obligatory," they were not of a uniform subjunctive type. (Some were volitional, others negated epistemics…etc…). In this task, the HSs were less accurate than the Spanish-dominant control group with expected subjunctive mood forms. Despite their lower accuracy, however, both the advanced HSs (86.2%) and the intermediate HSs (60.9%) again performed above chance.

Most relevant to the present study is the sentence conjunction task, which tested participants' interpretation of mood in three different constructions: adjectival relative clauses, temporal clauses with *cuando* ('when') and purpose clauses with *de manera que* ('so that'). For the purposes of this brief summary, I will focus on the adjectival relative clauses, given their special relevance to the present dissertation. Participants in the task read conjoined sentences, such as (20) below, and rated how "logical" they were, using a -2 (contradictory) to +2 (logical) scale. The logical (or contradictory) nature of a given sentence pair was dependent on the mood morphology of a verb within the conjoined

sentences. In (20), for example, the indicative mood morphology on the verb *tiene* ('have') makes the sentence "contradictory," given that (a) indicative mood implies that the speaker presupposes the existence of the book, yet (b) the speaker denies such presupposition at the end of the sentence by stating *no sé si hay uno* ('I don't know if there is one').

(20) *Necesito un libro de cuentos para niños **que tiene**$_{\mathbf{INDIC}}$ ilustraciones de Miró pero no sé si hay uno. (Contradictory)

(21) Necesito un libro de cuentos para niños **que tenga**$_{\mathbf{SUBJ}}$ ilustraciones de Miró pero no sé si hay uno. (Logical)

In (21), on the other hand, the subjunctive mood morphology on the verb *tenga* ('have') makes the sentence "logical," given that (a) subjunctive mood implies that the speaker does *not* presuppose the existence of the book and (b) the speaker expresses a matching lack of presupposition in the second half of the sentence. Participants who comprehend the mood morphology in adjectival relative clauses would be expected to rate logical sentences like (21) significantly higher than contradictory sentences such as (20).

Not surprisingly, the Spanish-dominant controls made this distinction quite clearly, providing average ratings of 1.75/2 for "logical" sentence pairs (like (21)) and -1.22/2 for "contradictory" sentence pairs (like (20)). The HSs, however, diverged from this pattern, providing significantly higher ratings for the "contradictory" sentence pairs. Nonetheless, the within-group results suggest that the HSs, at least at the advanced proficiency level, are sensitive to mood distinctions in adjectival relative clauses. The advanced HSs rated "logical" sentence pairs significantly higher (1.61/2) than "contradictory" sentence pairs (0.33/2), demonstrating qualitatively similar sensitivity to that of the Spanish-dominant

controls. The intermediate HSs, on the other hand, were not sensitive to mood distinctions in relative clauses, rating "logical" and "contradictory" sentence pairs equally.

Based on the fact that the HSs diverge from the controls in both productive and receptive assessments of mood knowledge, Montrul posits that "incomplete acquisition appears to affect their [HSs'] grammatical competence more globally, not just aspects of oral performance" (p. 265). For Montrul, therefore, HSs' divergence from controls suggests that they have different underlying representations of mood morphology, in line with the Representational Differences approach presented in Chapter 2 (e.g., Montrul, 2002, 2008).

It is certainly the case that the HSs in this study exhibit greater variability than the controls across all three experimental tasks. In the case of the intermediate HSs, who do not differentiate between subjunctive and indicative relative clauses, it seems quite likely that they have different representational knowledge of polarity subjunctive.

Nonetheless, it is problematic to assume that the HSs, especially at the advanced proficiency level, have "incomplete knowledge" of subjunctive. If they have "incompletely" acquired the features underlying subjunctive mood, what allows them to produce it so frequently? Similarly, if they have "incomplete" knowledge of mood, what allows the advanced HSs to rate "logical" subjunctive-marked relative clauses significantly higher than "contradictory" indicative-marked relative clauses? Montrul (2009) does not offer an explanation for how "incomplete" underlying knowledge can lead HSs to exhibit systematic within-group distinctions. In addition, Montrul's study does not address the issue of HS variability, as defined in Chapter 2. What factors lead individual HSs to alternate between target-like subjunctive forms and non-target indicative mood forms?

In a more recent study, van Osch & Sleeman (*forthcoming*) used an elicited production task to test simultaneous HSs' (n=17; average DELE proficiency = 42.4) production of intensional and polarity subjunctive mood. Participants read short contexts, each of which was followed by a sentence fragment that they finished out loud. The intensional condition tested participants' subjunctive production with *querer que* while the polarity conditions tested participants' subjunctive production in adjectival relative clauses and negated epistemics (e.g., *no creo que*…: 'I don't believe that…').

In both intensional and polarity conditions, the HSs produced significantly less subjunctive mood forms than a Spanish monolingual control group. With *querer que* participants produced subjunctive approximately 80% of the time, compared to nearly 100% for the controls. With non-presuppositional relative clauses, however, the HSs diverged even more sharply from the controls, producing subjunctive mood in only about 50% of all items. A within-group comparison revealed that the HSs did not differentiate between subjunctive and indicative in these clauses, suggesting a lack of polarity mood sensitivity. To account for the differential behavior of the HSs in intensional and polarity conditions, the authors reference the Interface Hypothesis (Sorace & Filiaci, 2006), which predicts increased bilingual optionality in so-called interface properties.

At the end of the paper, van Osch & Sleeman (*forthcoming*) compare the results from the elicited production task to results from an acceptability judgment task (AJT) completed by the same participants. Unlike in the elicited production task, the HSs in the AJT showed mood sensitivity in non-presuppositional relative clauses, a finding which the authors attribute to their (a) higher metalinguistic knowledge and (b) less frequent opportunities to use Spanish. Another possibility, unexplored in their paper, is that the HSs'

decreased accuracy in the production task is a function of on-line difficulty in the activation of mood features for *production* (Putnam & Sánchez, 2013).

Perez-Cortes (2016) used a series of innovative tasks to test HSs' production and comprehension of intensional and polarity subjunctive mood forms in Spanish. The intensional subjunctive condition in Perez-Cortes's study tested subjunctive with *querer que* while the polarity subjunctive condition tested subjunctive with *decir que* ('say that').

In the elicited (written) production task, participants read short contexts, each of which was followed by the presentation of an incomplete sentence on screen. Participants were asked to complete the sentence using a form of the verb provided to them in parentheses. In the intensional subjunctive condition, the advanced HSs produced subjunctive mood in 93.8% of items, compared to 62.7% for the intermediate HSs. Interestingly, the HSs' rate of subjunctive production was nearly identical with polarity subjunctive, where the advanced HSs produced subjunctive 93.7% of the time and the intermediate HSs once again produced it in 62.7% of items. In an Acceptability Judgment Task and a Truth Value Judgment Task, these two groups of HSs also exhibited *receptive* knowledge of intensional and polarity subjunctive mood forms.

The participants in Perez-Cortes's study, like the HSs in Montrul (2009) and van Osch & Sleeman (*forthcoming*), exhibit considerable variability in the production of subjunctive mood morphology, often alternating between subjunctive and indicative forms within the same condition. Perez-Cortes, following Putnam & Sánchez (2013), suggests that these mood alternations could be the result of "difficulties in the remapping of functional features (FFs) onto their corresponding morphological forms during production"

(p. 227). In other words, some of HSs' mood variability may be related to HL activation rather than representational differences.

A surprising finding of this study is that HSs perform equally well with intensional and polarity subjunctive mood forms, a result that Perez-Cortes attributes to the fact that both *querer* and *decir* are deontic predicates (e.g., volitionals and reported directives). Previous studies finding advantages for intensional subjunctive forms (e.g., van Osch & Sleeman, *forthcoming*; Pascual y Cabo, Rothman & Lingwall, 2012) had not investigated two forms within the same modality type, leaving open the possibility that the mood selection effects (intensional vs. polarity) were really an artifact of modality differences.

The final study which I will review is Giancaspro (under revision), which served as a pilot for the present dissertation project. In the study, I investigated 39 HSs' (26 advanced proficiency and 13 intermediate proficiency) oral production of intensional subjunctive (with *para que*) and polarity subjunctive (in adjectival relative clauses) mood forms.

The main instrument was the Contextualized Elicited Production Task (CEPT), loosely based on Pérez-Leroux (1998). Participants were asked to imagine themselves in a department store setting, where they were shopping. For each experimental item, participants read a three-sentence context (in Spanish) about a product (with a certain characteristic) that they are looking for. Then, they recorded themselves finishing incomplete sentences, using a form of a verb provided to them in parentheses, in order to communicate their needs to a clerk at the store. Examples of polarity subjunctive and intensional subjunctive items are provided in (22) and (23).

(22) Polarity Subjunctive Condition
Context: "You're in the school section. You need to buy a new calculator with four extra batteries. You find one without any extra batteries and another with only 2 extras."

You tell the clerk:
"Busco una calculadora que _____ (INCLUIR)…"
'I am looking for a calculator that _____ (INCLUDE)…"
Expected Mood: *incluya* 4 pilas extras…
'include-SUBJ 4 extra batteries…'

(23) Intensional Subjunctive Condition
Context: "You're in the music section. Your brother rarely plays music, but he needs a guitar for his music class. You like to buy him musical gifts."

You tell the clerk:
"Busco una guitarra *para que* mi hermano _____ (PRACTICAR)…"
'I'm looking for a guitar so that my brother _____ (PRACTICE)…"
Expected Mood: *practique más*…
'practice-subj more…'

The results of the study, compiled in Table 2, revealed that HSs exhibit variability in their production of both intensional and polarity subjunctive mood forms. Further analyses indicated that HSs' mood variability is strongly affected by both between-group and within-group factors. From a between-group standpoint, Spanish proficiency (and also, AofA Eng, though this is not presented here) significantly predicted the extent to which HSs exhibited variability. Not surprisingly, advanced HSs were significantly more likely to produce subjunctive mood than intermediate HSs. From a within-group standpoint, there were significant effects of both mood selection and verb frequency. Not only were HSs more likely to produce intensional subjunctive than polarity subjunctive, they were also more likely to produce subjunctive with frequent verbs.

Based on the advanced HSs' above-chance performance with both intensional and polarity subjunctive, I argued against the Representational Difference (e.g., Montrul, 2002,

2008, 2009) account of HS divergence, suggesting that HSs could not perform in such a systematic way without target-like mood representation. Based on the results of the Spanish-dominant controls, who exhibited no variability with subjunctive mood, I also argued against an Input Quality (e.g., Pires & Rothman, 2009) account for the HSs' divergence and variability. Considering that the HSs' likelihood of subjunctive mood production increased with frequent verbs, I argued that results best supported the Activation/Lexicalist approach (Putnam & Sánchez, 2013) to HS acquisition.

Table 2
Participants' probability of producing subjunctive

Group	Mood Selection	Verb Frequency	Logodds	SE	Odds	Probability	CI-Lower	CI-Upper
SDC (n=14)	Intensional	Frequent	4.026	1.247	56.04	98.2%	82.7%	99.8%
		Infrequent	4.144	1.244	63.05	98.4%	84.0%	99.9%
	Polarity	Frequent	4.103	1.243	60.52	98.4%	83.9%	99.9%
		Infrequent	4.118	1.247	61.44	98.4%	84.0%	99.9%
AdvHS (n=26)	Intensional	Frequent	3.725	.785	41.47	97.6%	89.5%	99.5%
		Infrequent	2.781	.709	16.14	94.2%	79.0%	98.6%
	Polarity	Frequent	1.368	.648	3.93	79.7%	50.2%	93.9%
		Infrequent	0.792	.635	2.21	68.8%	36.7%	89.4%
IntHS (n=13)	Intensional	Frequent	0.377	.770	1.46	59.3%	23.2%	87.6%
		Infrequent	-.007	.768	.990	49.8%	17.1%	82.7%
	Polarity	Frequent	-1.531	.805	.220	17.8%	4.1%	52.6%
		Infrequent	-2.752	.886	.06	6.0%	1.1%	27.5%

(from Giancaspro, under revision)

One key shortcoming of Giancaspro's study is that it lacked *receptive* tests of HSs' subjunctive mood knowledge, making it impossible to know whether the HSs who did not consistently produce subjunctive also did not comprehend it. Without such critical data, it is extremely difficult to distinguish between accounts for HS divergence and variability.

3.4. Chapter Summary

In Chapter 3, I have presented critical information about the property of interest in the present study: subjunctive mood morphology. At the beginning of Chapter 3, I described the different surface realizations of subjunctive mood in Spanish, showing that subjunctive mood is marked with a thematic vowel shift and, in some cases, a shift in the verbal root, as well. Then, I presented two subtypes of subjunctive mood: intensional subjunctive and polarity subjunctive, arguing that these two subtypes (a) have different syntactic representations and (b) may be stored differently in the mental lexicon. Finally, after outlining the learning task faced by Spanish speakers acquiring subjunctive mood in Spanish, I reviewed research on acquisition of intensional and polarity subjunctive mood by L1 Spanish speakers, early bilingual children and adult HSs.

In Chapter 4, I will present the methodology of the present study, which incorporates both productive and receptive experiments and, consequently, offers the opportunity to more comprehensively examine HSs' knowledge of intensional and polarity subjunctive mood forms in Spanish.

CHAPTER 4:
METHODOLOGY

4.1 Introduction

The goal of the present study, broadly speaking, is to examine HSs' knowledge of intensional and polarity subjunctive and indicative mood morphology in Spanish, as well as between- and within-group factors which condition such knowledge. Because task modality can play a significant role in HSs' performance (e.g., Montrul, Foote & Perpiñán, 2008), three experimental tasks are employed in hopes of providing participants with different, and differentially taxing, ways of demonstrating their sensitivity to intensional and polarity subjunctive forms in Spanish.

In this chapter, I will first present the RQs which guide the present study. Then, I will provide a detailed description of the four experimental groups, focusing on differences in age of acquisition of English (AofA Eng), Spanish and English proficiency, as well as usage of and exposure to Spanish with family and friends. Finally, I will outline the three experimental tasks, showcasing sample items and highlighting crucial elements of the methodological design which will have signficant importance during the discussion and presentation of the results in Chapter 5.

4.2 Research questions

As pointed out in Chapter 2, HL acquisition is characterized by great *inter*- and *intra*-individual variability in the production and comprehension of the HL. Not only is there immense variability in the grammatical knowledge of different HSs (e.g., Montrul, 2008, 2016), there is also great variability *within* individual HSs themselves (e.g., Giancaspro, under revision). While both inter- and intra-individual variability are

frequently observed in the HL literature, it is still not clear what extra-linguistic and linguistic factors underlie such variability and to what extent.

In the case of inter-individual variability, what extra-linguistic factors influence the likelihood that HSs will demonstrate target-like knowledge of grammatical properties, such as subjunctive mood morphology? By examining the roles of proficiency, AofA Eng and frequency of Spanish usage, the present study will shed light on the combined impact of these related, but distinct, variables on HSs' knowledge of subjunctive mood.

In the case of intra-individual variability, what linguistic factors cause an individual HS to demonstrate variable production and comprehension of a single grammatical form, such as mood morphology, either (a) in a given testing session or (b) across his/her lifetime (Silva-Corvalán, 1994)? By addressing the impact of structural priming, as well as lexical selection and frequency, on HSs' intra-individual variability, the present study will attempt to determine whether such HS variability is a result of representational differences in the HL syntax (e.g., lacking a particular syntactic feature) or lexicon (e.g., not accessing the syntactic feature in real time or, alternatively, not mapping a particular feature to certain lexical items). The answer to this question will have implications not just for heritage langage research but also for our understanding of the complex interaction of lexical and morphosyntactic knowledge in language acquisition more generally.

With these goals in mind, the following RQs guide the present study.

1. What is the nature of HSs' knowledge of lexically (intensional) and contextually (polarity) selected mood morphology in Spanish?

Based on many previous studies of HSs' knowledge of subjunctive mood (Giancaspro, under revision; Montrul, 2007, 2009; Montrul & Perpiñán, 2011; Pascual y Cabo et al., 2012; Perez-Cortes, 2016), it is hypothesized that HSs will diverge from Spanish-dominant controls in both the production and comprehension of subjunctive mood morphology. In oral production, HSs are expected to produce significantly less subjunctive mood than the Spanish-dominant controls in contexts where subjunctive is expected. In those instances, HSs are expected to produce underspecified, indicative mood forms. In acceptability and preference tasks, HSs are expected to (a) less strongly differentiate between subjunctive and indicative or (b) not differentiate between these forms at all.

It is critical, as embodied in RQ #1, for us to *describe* the characteristics' of HSs' sensitivity to subjunctive and indicative mood forms. Nonetheless, *describing* HSs' mood knowledge is only the first step. The remaining RQs, therefore, target factors which shape and *explain* HSs' sensitivity to mood.

2. What is the role of between-group factors on HSs' knowledge of subjunctive mood?

HSs are a notoriously heterogeneous group (Montrul, 2016a) who differ along a number of dimensions including AofA Eng, HL proficiency, birth order, and activation of the HL (Perez-Cortes, 2016), among many other demographic and social factors. Consequently, RQs 2a and 2b explore the extent to which a few of these between-group variables impact HSs' knowledge of mood.

2a. What is the impact of AofA Eng on participants' sensitivity to mood?

It is hypothesized that AofA Eng will correlate negatively with mood sensitivity, meaning that participants with later AofA Eng are expected to demonstrate the most productive and receptive sensitivity to mood in Spanish.

In the present study, this means that the Spanish-dominant controls (SDCs), who acquired English at age 13 or later, will show the greatest sensitivity to mood distinctions, followed in descending order by the late childhood immigrants (LCIs), who acquired English between the ages of 8 and 12, and, finally, the HSs, who acquired English before the age of 7. This prediction is motivated by previous findings from both HL and L1 attrition research.

Within HL research, many studies (e.g., Giancaspro, under revision; Pascual y Cabo & Gómez-Soler, 2015; though see Montrul & Sánchez-Walker, 2013 for an exception) have found that simultaneous HSs, exposed to English in the earliest stages of language development, demonstrate less productive and receptive sensitivity to properties of the HL than sequential HSs, who began learning English at school age. Similar differences have also been found between sequential HSs and later childhood immigrants, who began learning English in the US between age 8 and age 12 after having previously lived in a Spanish-speaking country. Montrul (2002), for example, found that later childhood immigrants performed more accurately than sequential HSs in the production and comprehension of aspectual morphology in Spanish.

L1 attrition research (e.g., Bylund, 2009; Schmid, 2012) has identified age 12 as a crucial inflection point beyond which exposure to an L2 is significantly less likely to result in attrition to and variability in the L1/HL. This observation supports the prediction that the SDCs in the present study will show less attrition of, and therefore greater sensitivity

to, mood distinctions in Spanish than both the LCIs and the HSs. The age at which bilinguals begin acquiring a second, societally dominant language clearly has an enormous impact on acquisition and maintenance of linguistic properties of the L1, especially when we draw broad comparisons between those who began acquiring the L2 before vs. after puberty. Nonetheless, when exposure to the L2 occurs before puberty, the age of acquisition effect diminishes (Bylund, 2009) and linguistic factors such as HL proficiency take on a greater predictive role. RQs 2a and 2b examine the role of these factors on HSs' mood sensitivity.

2b. What is the impact of Spanish proficiency on HSs' sensitivity to mood?

It is hypothesized, based on many previous studies of HSs and verbal morphology (e.g., Giancaspro, under revision; Montrul, 2002, 2007, 2009; Montrul & Perpiñán, 2011; Perez-Cortes, 2016) that HSs with higher HL proficiency will demonstrate more productive and receptive sensitivity to mood distinctions than HSs with lower proficiency in the HL.

HL proficiency, whether measured via self-ratings or standardized assessments, has consistently predicted HSs' performance with different linguistic properties. Nonetheless, it is important to acknowledge that proficiency is a notoriously difficult concept to define (Austin, Blume & Sánchez, 2015) and to disentangle from other potentially important variables including, but not limited to, relative linguistic dominance (Montrul, 2016b) as well as the use of (e.g., Perez-Cortes, 2016), exposure to (e.g., Grüter & Paradis, 2015) and societal status of (e.g., Valenzuela et al, 2015) the HL.

Giancaspro (under revision) divides his HS participants into advanced and intermediate proficiency on the basis of the DELE proficiency test but notes a strong

relationship between AofA Eng and HL proficiency: while most simultaneous HSs scored in the intermediate proficiency range, the vast majority of sequential HSs scored in the advanced range. This strong correlation between AofA Eng and proficiency may be at least partly attributable to differences in the status of the HL for simultaneous and sequential HSs, respectively. (In other words, HSs who are not exposed to English until later may be more likely to live in communities where Spanish has a higher societal status and is used more.) HL proficiency also appears to be correlated with use of the HL. Perez-Cortes (2016) reports strong positive correlations between Spanish proficiency, as measured by the DELE, and participants' self-reported Spanish usage.

3. What is the role of within-group factors on HSs' sensitivity to subjunctive mood?

RQs 2a and 2b take as a starting point the idea that HSs, even when matched for age of acquisition of English, have vastly different backgrounds which can impact their knowledge of certain linguistic properties. In RQs 3a, 3b and 3c we shift our focus to differences in the linguistic property, subjunctive mood, and within-group factors which condition its production and comprehension by HSs.

3a. What is the role of mood selection type?

It is hypothesized that HSs will exhibit less divergence/variability with intensional subjunctive than with polarity subjunctive for a few different reasons. First, as highlighted in Chapter 3, intensional subjunctive with *para que* is earlier acquired than polarity subjunctive in relative clauses. Previous research suggests that HSs exhibit less variability with earlier acquired properties, e.g., aspectual morphology vs. mood morphology in

Spanish (Montrul, 2009; Montrul & Perpiñán, 2011) or continuous vs. discontinous morphological negation in Arabic (Albirini & Benmamoun, 2015). One possible reason for this trend is that earlier acquired properties tend to be somehow less cognitively "complex"[16]. Alternatively, early acquired forms may just have more time to stabilize before the societally-dominant language begins to exert its influence on the HL.

The second reason to hypothesize that HSs will exhibit less divergence/variability with intensional subjunctive mood has to do with the way in which intensional and polarity subjunctive forms may be stored in the mental lexicon. Jackendoff (2007) argues that the human lexicon includes not just words but also bigger units known as treelets. Treelets, as discussed in Chapter 3, are larger syntactic templates (e.g., DPs headed by *the*) stored in long-term memory, where they can facilitate anticipatory processing. The English determiner *the* almost always precedes a NP. Therefore, Jackendoff argues that English speakers store *the* in a treelet alongside its complementary NP structure, which presumably allows for faster production and comprehension of *the*-headed DPs.

Similar to the English determiner *the*, the Spanish complementizer *para que* obligatory selects for complements with the uninterpretable *u*W mood feature (Kempchinsky, 2009). It is feasible, therefore, to suggest that HSs of Spanish store *para que* as a treelet alongside its complementary structure in their mental lexicons, making it

---

[16] Complexity is often poorly defined in bilingualism research (See Scontras et al., 2015 for discussion). Nonetheless, it is still possible to draw rough but principled distinctions between more and less complex instantiations of a given property (e.g., Culicover, 2013). Albirini & Benmamoun (2015), for example, argue that discontinuous negation in Arabic is more "complex" than continuous negation because, unlike continuous negation, it involves syntactic movement. Sorace (2011), in a similar vein, argues that linguistic properties which involve both core grammatical knowledge and discourse/pragmatic sensitivity are inherently more "complex" than purely grammatical properties with no such interface. Pérez-Leroux (1998) argues that properties which require Theory of Mind, e.g., polarity subjunctive mood, are inherently most complex. In this paper, I use the term "complex" in a comparative sense only, that is to say, when analyzing the extent to which two instantiations of a given property involve syntactic movement, integration of discourse/pragmatic information or differential storage in and accessibility from the lexicon.

relatively easier to access and produce. The interpretable W feature that triggers polarity subjunctive mood, on the other hand, is associated with a lack of presuppositional context rather than any triggering lexical item. Consequently, HSs cannot store the W feature as a treelet alongside co-occurring lexical items, making it more difficult for them to access and produce this interpretable W feature in real time.

The third and final reason to hypothesize that HSs will exhibit less variability with intensional subjunctive mood forms comes from previous research. Though Perez-Cortes (2016) did not find differences between HSs' knowledge of polarity and intensional subjunctive forms, Giancaspro (under revision) tested the same properties examined in the present dissertation and found that HSs exhibited significantly less variability with intensional subjunctive mood forms.

3b. What is the role of priming status on HSs' productive sensitivity to mood?

It is hypothesized that HSs will produce more subjunctive forms after exposure to structural primes (Pickering & Ferreira, 2008) than in baseline, non-prime conditions. The motivation for this hypothesis is as follows. If HSs have some underlying knowledge of the features which trigger subjunctive mood, then exposure to these features immediately prior to production should make it easier for them to access these features and produce subjunctive mood in a target-like fashion. As pointed out by Sturt, Keller & Dubey (2010), "the processing of a linguistic unit is facilitated by the recent processing of a linguistic unit with an equivalent syntactic form" (p. 333).

3c. What is the impact of verb frequency on HSs' sensitivity to mood?

It is hypothesized that HSs will show greater productive and receptive sensitivity to mood with more frequent lexical items. This prediction is based on frequency effects in L2 acquisition, as well as preliminary evidence—both experimental and non-experimental—from HL acquisition.

Lexical frequency effects have been extensively documented in second language acquisition research, as noted by Ellis (2002), Ellis & Collins (2009), and Wulff, Ellis, Römer, Bardovi-Harlig & Leblanc (2009), among many others. Most relevant for the present study is the well-established effect of token frequency (how often a given lexical form appears in the input) on second language learners' acquisitional outcomes. Wulff et al point out that high frequency lexical items, by definition, are "more likely to be experienced" by second language learners (p. 366) than low frequency lexical items, making high frequency lexical items both easier to access (representationally) and more automatized for productive use. Frequency effects are not, however, limited to production. Ellis (2002), in his review of frequency effects in L2 acquisition research, reveals that L2 learners are also better at *recognizing* high-frequency words in auditory input.

If these well-documented frequency effects extend to HSs, then it is hypothesized that HSs in the present dissertation will be better at both producing and recognizing subjunctive mood morphology on higher frequency verbs. Strengthening the likelihood of this possibility is previous research from Dorian (1981), Gal (1989) and Giancaspro (under revision), as well as theoretical argumentation from Albirini (2014), Dorian (1981), Gal (1989), Polinsky (2006) and Putnam & Sánchez (2013), all of whom have discussed frequency effects in HL acquisition.

Dorian (1981), in a seminal study of HSs of Gaelic, reports that her participants are more accurate with frequent, as opposed to infrequent, conditional morphology. Gal (1989) presents similar findings in a study of Hungarian HSs, who utilize causative morphology most accurately with frequent verb forms. However, because both of these studies are largely non-experimental in nature, it is hard to more specifically quantify the exact impact of lexical frequency on HSs' knowledge and production of verbal morphology.

Giancaspro (under revision) tested HSs' production of subjunctive mood morphology and found that HSs, in an elicited production experiment, were significantly more likely to produce subjunctive mood morphology on frequent, as opposed to infrequent lexical items. To my knowledge, this is the only HL study to test lexical frequency effects experimentally.

Theoretically, there are a few reasons to expect frequency effects both in HL acquisition more broadly, as well as in the present dissertation project. One reason to expect HSs to perform more accurately with frequent lexical items is chunking (e.g., Polinsky, 2006). It is not hard to imagine that HSs could memorize certain, particularly frequent chunks (e.g., *que* sea_{SUBJ}, 'that is') without having abstract knowledge of the morphosyntactic features underlying those structures. If this were the case, we would expect to see strong frequency effects in the statistical analysis.

A second potential reason to expect lexical frequency effects is the underlying separation between lexical items and functional features (e.g., Adger & Smith, 2005, 2010). If it is possible to acquire a functional feature (e.g., the *u*W feature involved in intensional subjunctive) without necessarily knowing how that feature is instantiated on all lexical items (e.g., knowing that the subjunctive mood form of *observar* is *observe*), then HSs

could theoretically maintain knowledge of an abstract feature without knowing all of its lexical instantiations, resulting in the curious patterns discussed by Albirini (2014), Dorian (1981), Gal (1989) and Giancaspro (under revision).

Gal (1989) argues that HSs of Hungarian maintain their "combinatorial ability" (p. 327) to generate morphological causatives, though they are mostly able to use that ability with more frequent and frequently activated verbs. Similarly, Dorian (1981) argues that HSs of Gaelic maintain knowledge of abstract gender (and presumably gender agreement), though they apply this abstract knowledge to a "smaller number of nouns" (p. 129) than dominant speakers of Gaelic. In both cases, as well as in Giancaspro's study of subjunctive mood, HSs' featural knowledge is simply expressed on a smaller subset of lexical items, possibly due to reduced activation or use of those items (Putnam & Sánchez, 2013) in HL production and comprehension.

In the next section, I outline the methods and procedures which were employed in order to effectively test the RQs presented in Section 4.2.

4.3 Participants

All participants in the present study (n=81) are native speakers of Spanish living in the US. Because all participants were native speakers of Spanish, participants were first divided into groups according to their AofA Eng. The HSs (n=42) began acquiring English in the United States at or before age 6[17]. Because Spanish proficiency has been shown to

[17] It is important to note that there is no consensus as to when childhood bilinguals are no longer considered heritage speakers (e.g., Montrul, 2016a; Rothman, 2009). What is the justification, therefore, for excluding (from the HS groups) bilinguals who learned English between the ages of 8 and 12? Bilinguals who are exposed to the dominant language in later childhood (e.g., age 8+) tend to pattern differently from bilinguals with earlier exposure to the societally-dominant language (Lee, 2011; Montrul, 2002), e.g., demonstrating greater accuracy and sensitivity to properties of the HL. While there are also differences in the HL knowledge of so-called simultaneous (e.g., English from birth) and sequential (e.g., English at age 5) bilinguals (e.g., Pascual y Cabo & Gómez Soler, 2015), differences between these groups are less pronounced (Montrul &

play a significant role in HSs' knowledge of grammatical properties, the HSs were further subdivided into advanced (n=22) and intermediate (n=20) proficiency groups according to their performance on the fifty-question DELE proficiency test (Duffield & White, 1999; Bruhn de Garavito, 2002; Montrul & Slabakova, 2003; *inter alia*). The Spanish-dominant controls (SDCs) (n=20), who began acquiring English at age 13 or later, served as the 'baseline' against which to compare the HSs. Finally, the late childhood immigrants (LCIs) (n=19), who began acquiring English in the United States between the ages of 8 and 12, served as an intermediary control group[18].

This particular combination of experimental groups allows for a detailed analysis of the different between-group factors which impact production and comprehension of subjunctive mood morphology. By testing HSs (AofA Eng: birth to age 6), late childhood immigrants (LCIs; AofA Eng: age 8 to age 12), and Spanish-dominant controls (SDCs; AofA Eng: age 13 or later), the researcher allows for a gradient assessment of how relative increases in AofA Eng affect bilinguals' variability in subjunctive mood production and comprehension. By splitting the HSs into two proficiency groups, the researcher evaluates the role of HL proficiency, independent of AofA Eng on HSs' subjunctive knowledge.

Below, I provide more information about participants in each of these four groups.

4.3.1 Advanced HSs (AdvHSs)

Sánchez-Walker, 2013) than differences between earlier and later childhood bilinguals. By drawing the line at age six (instead of say, 8 or 9), I avoid grouping later childhood bilinguals, who are likely to exhibit much higher accuracy in the HL, alongside qualitatively distinct earlier childhood bilinguals.

[18] See Montrul (2002) and Lee (2011) for other examples of intermediary childhood immigrant groups.

Of the 22 AdvHSs, 20 (90.9%) were born in the US[19] and 2 (9.1%) moved to the US in early childhood (age 4 and age 6, respectively)[20]. All AdvHSs were second generation, that is to say, their parents immigrated to the US, and all AdvHSs scored 40 or higher on the DELE proficiency test. At the time of testing, most AdvHSs were in their early 20s (Mean age: 21.22 years, SD: 2.60, Range: 18-31). The AdvHSs' parents immigrated from Puerto Rico (4.5%), in addition to Spanish-speaking countries including Mexico (22.7%), Ecuador (18.2%), and Peru (9.1%), as well as Argentina (4.5%), Chile (4.5%), Colombia (4.5%), the Dominican Republic (4.5%), El Salvador (4.5%), Guatemala (4.5%), Honduras (4.5%) and Spain (4.5%). Only two of the AdvHSs (9.1%) had parents who came from two different Spanish-speaking countries. All participants were undergraduate or graduate students at the time of completing the experiments.

All AdvHSs rated their overall English and Spanish proficiency on a ten-point scale from 1 (beginner) to 10 (native-like). 14 of the 22 AdvHSs (63.7%) rated their English higher than their Spanish while eight AdvHSs (36.3%) rated themselves as equally in each language. (No participants rated themselves as Spanish-dominant.) As a group, the AdvHSs' average self-rated English proficiency (M = 9.82, SD = 0.66) was significantly higher than their self-rated Spanish proficiency (M = 8.41, SD = 1.37), t (21) = 4.830, $p <$

---

[19] One of the AdvHSs who was born in the US moved to the Dominican Republic shortly after birth before returning to the US at age 5.

[20] As Montrul (2002) points out, HSs who lived abroad during childhood "appear to have a more stable grammar in Spanish" (p. 59) than HSs who are born in the United States. Despite this difference, I leave these two participants in the AdvHS group for two main reasons. First, both of these participants immigrated at ages 4 and 6, before receiving much formal instruction in Spanish. Second, it is unlikely that these HSs' Spanish proficiency and dominance upon immigrating differed significantly from the proficiency and dominance of HSs who were born in the US and lived with functionally monolingual Spanish-speaking parents until starting primary school.

.001, d = 1.13. Nonetheless, it is worth highlighting that the AdvHSs rated themselves quite highly (M = 8.41) in their weaker language, Spanish.

4.3.2 Intermediate HSs (IntHSs)

Of the 20 IntHSs, 17 (85%) were born in the US and 3 (15%) moved to the US in early childhood (ages 2, 2, and 5 respectively). 18 of the HSs (90%) were second-generation HSs while 2 of the HSs (10%) were third-generation HSs (e.g., their grandparents immigrated to the US). All of the IntHSs scored between 30 and 39 on the DELE. At the time of testing, the IntHSs were mostly in their late teens and early 20's (Mean age: 19.80 years, SD: 1.06, Range: 18-22). The IntHSs' parents immigrated from Colombia (20%), Ecuador (15%), Mexico (15%) as well as the Dominican Republic (5%), Guatemala (5%) and Peru (5%). Five of the IntHSs (25%) came from so-called "mixed" families where their parents immigrated from two different Spanish-speaking countries. All participants were undergraduate or graduate students at the time of testing.

19 of the 20 IntHSs (95%) rated their English proficiency higher than their Spanish proficiency while one of the IntHSs (5%) rated their English and Spanish proficiencies equally. As a group, the IntHSs' average self-rated English proficiency (M = 9.55, SD = 1.19) was significantly higher than their Spanish proficiency (M = 6.75, SD = 1.77), t (19) = 7.782, $p < .001$, d = 1.74. The large effect size for the IntHSs' self-rating comparison (d = 1.74) suggests that they are more strongly English-dominant than the AdvHSs.

4.3.3 Late Childhood Immigrants (LCIs)

Of the 19 LCIs, 15 (78.9%) were born in a Spanish-speaking country while 4 (21.1%) were born in the US. The four LCIs who were born in the US moved to a Spanish-

speaking country early in childhood (e.g., always at or before age 4) before returning to the US between the ages of 8 and 12[21]. All LCIs scored between 36 and 47 on the DELE[22]. At the time of testing, the LCIs were approximately the same age as the HSs (Mean age: 20.73 years, SD = 2.77, Range = 18-29). The LCIs immigrated (or immigrated *back*, in the case of the four LCIs born in the US) to the mainland US from Colombia (21.1%), Peru (21.1%), the Dominican Republic (15.8%), Ecuador (10.5%), Honduras (10.5%), Argentina (5.3%) and Puerto Rico (5.3%). Two of the LCIs (10.5%) came from "mixed" families where only one of the parents was born in a Spanish-speaking country. On average, the LCIs had been living in the US for an average of 10.45 years at the time of the experiments. All participants were undergraduate or graduate students at Rutgers at the time of testing.

Unlike the HSs, almost all of whom were English-dominant, the LCIs exhibited a greater range of relative language dominance. Eight of the LCIs (42.1%) rated their Spanish proficiency higher than their English proficiency while only two LCIs (10.5%) rated their English proficiency higher. The remaining nine LCIs (47.3%) rated themselves equally in English and Spanish. Overall, the LCIs' average self-rated Spanish proficiency (M = 9.58, SD = 0.77) was significantly higher than their self-rated English proficiency (M = 8.89, SD = 1.05), $t(18) = 2.387$, $p < .05$, $d = 0.56$. Nonetheless, the medium effect size of this statistical comparison ($d = 0.56$, cf. $d = 1.13$ for AdvHSs, 1.74 for IntHSs) suggests that

[21] It is possible, or perhaps even likely, that these participants learned some English before moving to a Spanish-speaking country. However, it is also likely that these participants' brief time living in the US was spent in Spanish-dominant households. All four of these participants reported that their parents used Spanish with them at least 85% of the time. In addition, none of these participants rated their parents' English proficiency greater than 6 out of 10.

[22] Of the 19 LCIs, 17 (89.5%) scored between 40 and 50, which is the standard "advanced" range on the DELE. The only two participants who did not score in this range were left in the LCI group.

the LCIs are, at least according to self-ratings, the most balanced of the three experimental (e.g., non-control) groups.

4.3.4 Spanish-dominant controls (SDCs)

All 20 of the SDCs were born in a Spanish-speaking country and lived there until moving to the US at age 13 or later (Mean age of arrival: 22.35, SD = 6.35, Range = 13-35). All SDCs scored between 40 and 48 on the DELE proficiency test. At the time of testing, the SDCs were slightly older than the HSs and LCIs (Mean age: 30.37 years, SD = 7.25, Range = 22-50) and had been living in the US, on average, for 8.01 years. SDCs came from Puerto Rico (5%) as well as Spain (20%), the Dominican Republic (15%), Mexico (10%), Ecuador (10%), Peru (10%), Costa Rica (10%), Bolivia (5%), Colombia (5%) and El Salvador (5%). All participants, with the exception of two Rutgers employees, were undergraduate or graduate students at the time of testing.

18 of the 20 SDCs (90%) rated their Spanish significantly higher than their English while two (10%) rated themselves as equally proficient in Spanish and English. Overall, the SDCs' average self-rated Spanish proficiency[23] (M = 9.95, SD = 0.22) was significantly higher than their self-rated English proficiency (M = 7.70, SD = 1.34), t (19) = 6.957, $p <$ .001, d = 1.56. Judging by the high effect size (d =1.56, cf. d = 0.56 for LCIs, d = 1.13 for AdvHSs and d = 1.74 for IntHSs) of the statistical comparison of their self-ratings, we can conclude that the SDCs are, relatively speaking, less balanced in their self-rated language proficiency than the AdvHSs and LCIs and slightly more balanced than the IntHSs. Nonetheless, the SDCs' high English self-rating (M = 7.70) shows that this group is

[23] Because all 20 of the SDCs rated their Spanish as 10/10, there was no variance in the data. In order to run the relevant paired-samples t-tests, therefore, I changed one SDC's Spanish SR from 10 to 9.

bilingual and, therefore, a fair baseline against which to compare the bilingual experimental groups. (See Pascual y Cabo & Rothman, 2012 for further discussion.)

4.3.5 Between-group comparisons

4.3.5.1 DELE

A one-way ANOVA was conducted to determine differences in the DELE scores of the four experimental groups. There were no outliers, as assessed by a visual examination of the box plots (Figure 1). Data was normally distributed for the IntHSs, Shapiro-Wilk test ($p > .05$) but not for the SDCs, LCIs and AdvHSs ($p < .05$). There was heterogeneity of variance according to Levene's test of homogeneity of variances ($p < .05$) so the Welch ANOVA was used for between-group comparisons. DELE scores were significantly different for the different participant groups, $F(3, 40.173) = 60.464$, $p < .001$. Games-Howell post-hoc analysis revealed significant differences in DELE scores between the SDCs (M = 45.60, SD = 2.62) and the AdvHSs (M = 42.41, SD = 1.74), $p < .001$, as well as between the SDCs and the IntHSs (M = 34.80, SD = 2.69), $p < .001$, and the LCIs (M = 44.00, SD = 3.61) and the IntHSs, $p < .001$. Curiously, the LCIs' DELE scores were not significantly different from the scores of the SDCs ($p > .4$) or the AdvHSs ($p > .3$).

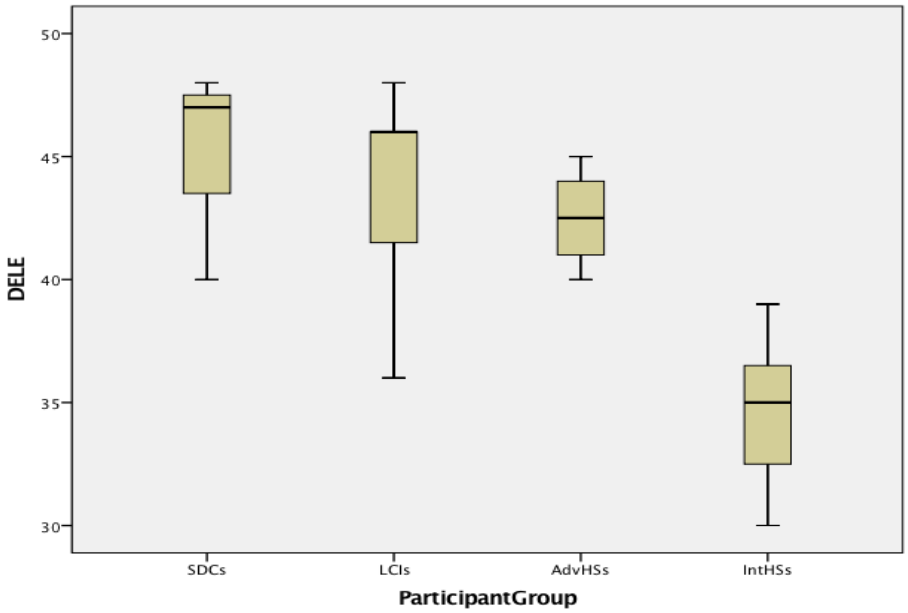

Figure 1: DELE scores by experimental group

4.3.5.2 English self-ratings

A one-way ANOVA was conducted to probe for group differences in average English self-ratings (SR's). A visual inspection of the box plot (Figure 2) revealed a total of ten outliers (out of 81 total participants). Data was normally distributed for the SDCs, Shapiro-Wilk test ($p > .05$) but not normally distributed for the other experimental groups ($p < .05$). There was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p > .05$). English SR's were significantly different across the experimental groups, $F(3,77) = 15.584$, $p < .001$, $\varpi^2 = 0.35$. Bonferroni post-hoc tests revealed that the AdvHSs' (M = 9.82, SD = 0.66) and the IntHSs' (M = 9.55, SD = 1.19) English SR's were significantly higher than the SR's for the SDCs (M = 7.70, SD = 1.34), $p < .001$ for both comparisons. The LCIs' (M = 8.89, SD = 1.05) SR's in English were also significantly higher than the SDCs' SR's. The AdvHSs' SR's were marginally higher ($p = .048$) than the

LCIs' ratings, but there were no pairwise differences between the LCIs and the IntHSs ($p > .3$) or the IntHSs and the AdvHSs ($p > .9$).

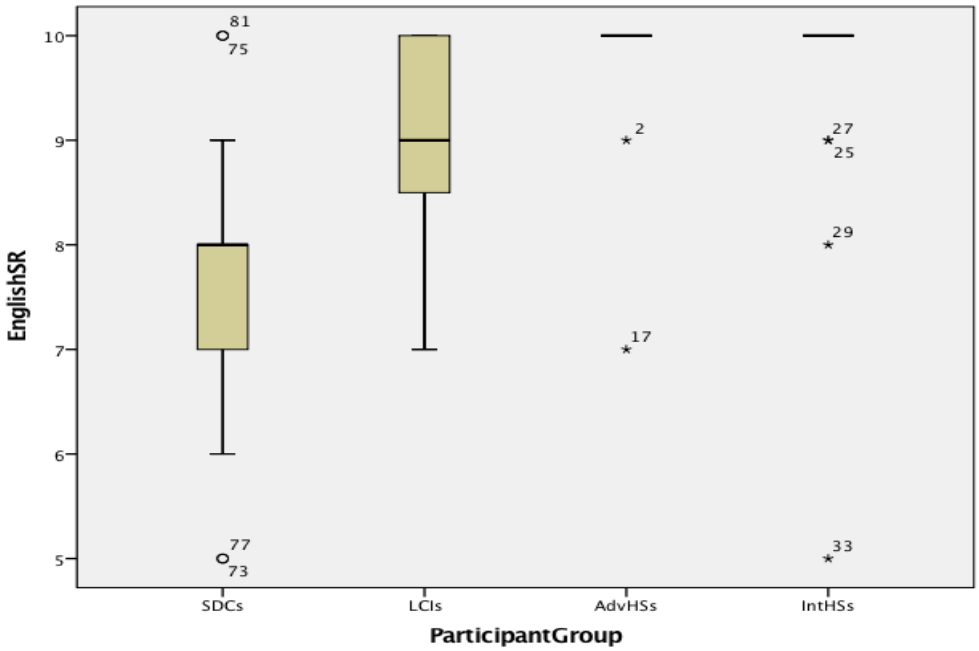

Figure 2: English self-ratings by experimental group

A Spearman's rank-order correlation was run to assess the relationship between the HSs' English SR's and DELE scores (Section 4.3.5.1). Visual inspection of the scatterplot revealed a monotonic relationship between the two variables. Results of the analysis reveal a very weak correlation between English SR's and DELE scores, $r_s = .021$, $p > .8$. That is to say, HSs' Spanish proficiency, as measured by the DELE, is not correlated with their English SR's. This finding is expected, given that HSs almost uniformly develop high proficiency in the societal language (e.g., Portes & Rumbaut, 2006) but only sometimes become highly proficient in the HL.

4.3.5.3 Spanish self-ratings

A one-way ANOVA was conducted to explore differences in Spanish SR's across the experimental groups. A visual examination of the box plot (Figure 3) revealed four outliers, all from the SDCs (1) and the LCIs (3). Data was normally distributed for the IntHSs (Shapiro-Wilk test: $p > .05$) but not for the SDCs, LCIs or AdvHSs ($p < .05$).

There was heterogeneity of variance, according to Levene's test of homogeneity of variance ($p < .05$), so the Welch ANOVA was used for the remainder of the analysis. The Welch ANOVA revealed significant differences in Spanish SR's across the experimental groups, $F(3, 34.512) = 29.466$, $p < .001$.

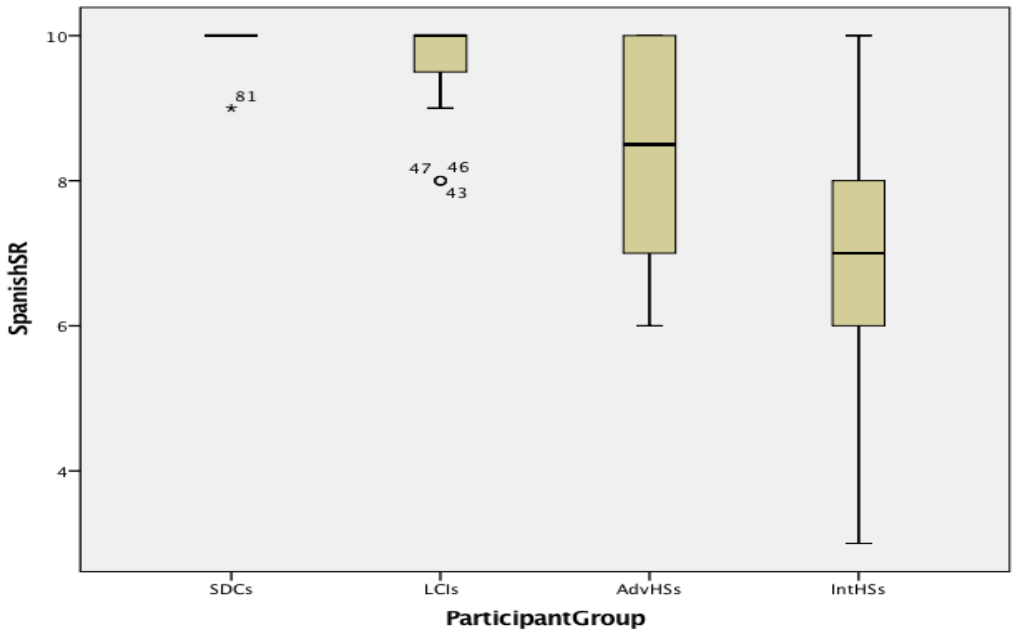

Figure 3: Spanish self-ratings by experimental group

Games-Howell post-hoc comparisons showed that the SDCs' Spanish SR's (M = 10.00, SD = 0.00)[24] were significantly higher than the ratings of the AdvHSs (M = 8.41, SD = 1.37; $p < .001$) and the IntHSs (M = 6.79, SD = 1.81; $p < .001$) but not the LCIs (M = 9.58, SD = 0.77; $p > .2$). The LCIs' Spanish SR's were higher than the ratings of the AdvHSs ($p < .01$) and the IntHSs ($p < .001$). Finally, the AdvHSs Spanish SR's were significantly higher than the ratings of the IntHSs, $p < .01$.

A Spearman's rank-order correlation was run to assess the relationship between the HSs' Spanish SR's and DELE scores (Section 4.3.5.1). Visual inspection of the scatterplot revealed a monotonic relationship between the two variables. Unlike in the case of the HSs' English SR's, which were not at all correlated with Spanish proficiency, the results of this analysis revealed a strong correlation between Spanish SR's and DELE scores, $r_s = .498$, $p < .01$. In general, HSs with higher DELE scores tended to rate their Spanish proficiency higher than HSs with lower DELE scores.

4.3.5.4 Spanish usage and exposure

Participants were asked to estimate what percentage of the time they used Spanish when conversing with their mothers, their fathers and their university friends. Participants' estimated usage percentages were averaged together to calculate an overall estimate of the extent to which participants in each group spoke Spanish with parents and friends.

As shown in Figure 4, the SDCs ($M_{\text{Spanish to Mom}} = 93.82$, SD = 9.11; $M_{\text{Spanish to Dad}} = 99.12$, SD = 2.64) and LCIs ($M_{\text{Spanish to Mom}} = 91.47$, SD = 10.20; $M_{\text{Spanish to Dad}} = 94.54$, SD = 7.43) used almost exclusively Spanish when speaking with parents while the AdvHSs

[24] All SDCs gave themselves a 10 self-rating in Spanish, meaning that there was no variation in their Spanish self-ratings. In order to run the Welch ANOVA, therefore, I changed one of the SDCs' Spanish self-ratings to a 9, as is reflected in Figure 3. Nothing critical hinges on this statistical decision.

($M_{Spanish\ to\ Mom}$ = 69.48, SD = 26.62; $M_{Spanish\ to\ Dad}$ = 67.52, SD = 28.91) and IntHSs ($M_{Spanish\ to\ Mom}$ = 55.37, SD = 33.77; $M_{Spanish\ to\ Dad}$ = 52.32, SD = 38.95) used a lot of, if also considerably less, Spanish with their parents.

The high standard deviations for the HS groups reveal the large variance in HSs' Spanish usage with parents. With university friends, the SDCs (M = 52.50, SD = 32.67) and LCIs (M = 30.89, SD = 25.43) use Spanish somewhat regularly while the AdvHSs (M = 17.00, SD = 16.48) and the IntHSs (M = 12.35, SD = 17.89) report using Spanish sparingly. The LCIs' higher usage of Spanish, as compared to the HSs, may be due to an increased likelihood for late childhood immigrants to forge social connections in their L1 (Jia & Aaronson, 2003; Oh & Fuligni, 2010). Thus far, I have argued that both the SDCs and LCIs are truly comparable bilingual control groups. The fact that both of these groups opt to use English equally, or slightly more than, Spanish with university friends is another indirect indicator of their relatively high English proficiency.

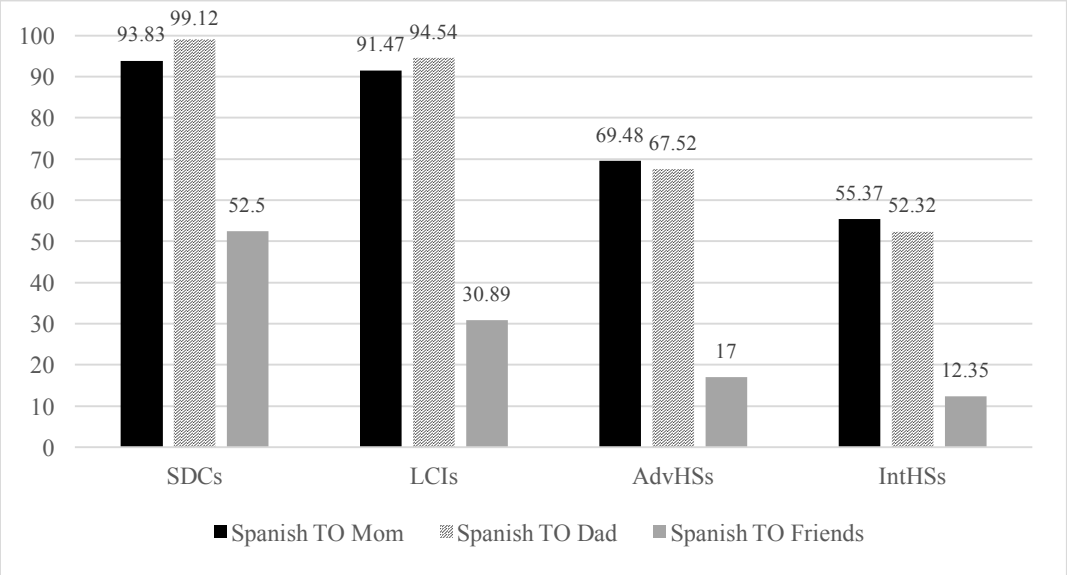

Figure 4: Participants' reported Spanish use with parents, friends

Because HSs often exhibit asymmetries in their Spanish usage and exposure, participants were also asked to estimate the percentage of the time in which their parents spoke Spanish to them. Not surprisingly, as shown in Figure 5, the SDCs ($M_{\text{Spanish from Mom}}$ = 97.06, SD = 12.13; $M_{\text{Spanish from Dad}}$ = 99.94, SD = 0.24) and the LCIs ($M_{\text{Spanish from Mom}}$ = 97.47, SD = 5.81; $M_{\text{Spanish from Dad}}$ = 96.15, SD = 7.40) reported that their parents speak to them almost exclusively in Spanish. The AdvHSs ($M_{\text{Spanish from Mom}}$ = 84.81, SD = 21.49; $M_{\text{Spanish from Dad}}$ = 77.43, SD = 27.67) and the IntHSs ($M_{\text{Spanish from Mom}}$ = 73.74, SD = 33.82; $M_{\text{Spanish from Dad}}$ = 68.89, SD = 38.99) reported that their parents, as well, spoke mostly Spanish to them, though not to the same extent as the parents of the SDCs and LCIs.

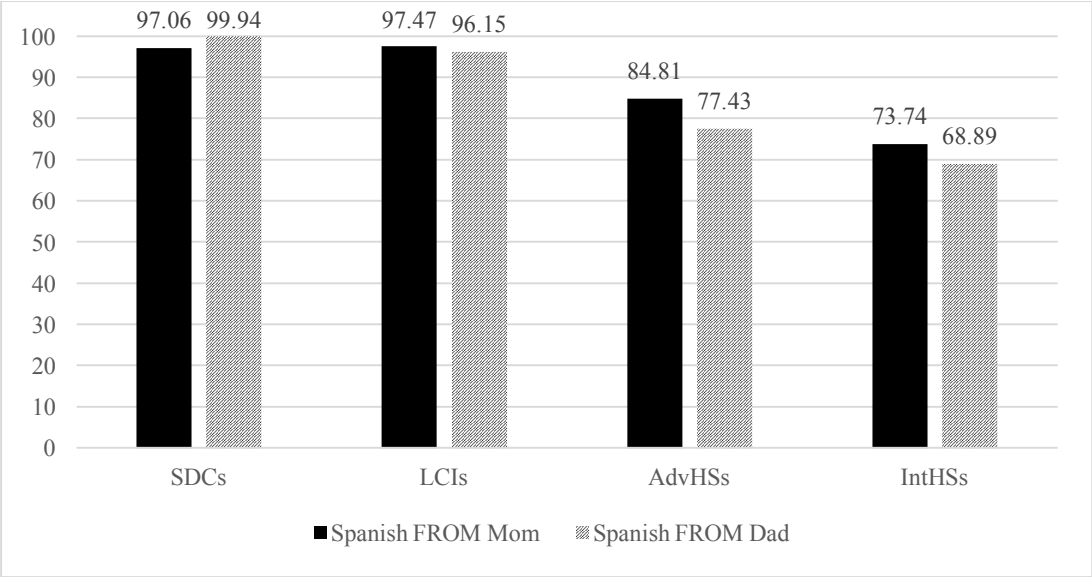

Figure 5: Participants' reported exposure to Spanish *from* parents

Figures 4 and 5, considered together, reveal that HSs use mostly Spanish when speaking with their parents, who also use mostly Spanish when speaking to them. As has been reported in many previous studies of HSs, however, the HSs report using more English with their parents than their parents use with them, a finding that may be indicative of shifting language dominance.

4.3.5.5 Summary of between-group comparisons

Not surprisingly, the SDCs differed from the HS groups in all proficiency, language usage and language exposure measures (Table 3). The SDCs were significantly more proficient in Spanish, both in self-ratings and in the DELE examination, than the HSs. However, despite reporting relatively high proficiency in English (M = 7.70/10) and high usage of English with friends at Rutgers (M = 47.50), the SDCs' English self-ratings were significantly lower than the ratings of the HSs. Finally, the SDCs reported using and hearing Spanish significantly more than the HSs both with parents as well as with friends at Rutgers. In summary, the SDCs are different from the HSs in terms of AofA Eng and Spanish proficiency. Nonetheless, their English proficiency is sufficient enough for them to be a maximally comparable bilingual control group.

Table 3.
Summary of participant groups

Participant Group	Age at Time of Testing	AofA English	DELE Proficiency	Spanish Self-Rating	English Self-Rating
AdvHSs (n=22)	Mean: 21.22 SD: 2.60	Age 6 or earlier	Mean: 42.41 SD: 1.74	Mean: 8.41 SD: 1.37	Mean: 9.82 SD: 0.66
IntHSs (n=20)	Mean: 19.80 SD: 1.06	Age 6 or earlier	Mean: 34.80 SD: 2.69	Mean: 6.75 SD: 1.77	Mean: 9.55 SD: 1.19
LCIs (n=19)	Mean: 20.73 SD: 2.77	Between Age 8 and Age 12	Mean: 44.00 SD: 3.61	Mean: 9.58 SD: 0.77	Mean: 8.89 SD: 1.15
SDCs (n=20)	Mean: 30.37 SD: 7.25	Age 13 or later	Mean: 45.60 SD: 2.62	Mean: 9.95 SD: 0.22	Mean: 7.70 SD: 1.34

Within the HS groups, there were no clear differences in self-rated English proficiency, as expected. Nonetheless, the AdvHSs and the IntHSs differed considerably in DELE proficiency, the factor which divided them into separate groups, as well as self-rated Spanish proficiency. Both HS groups were dominant in English, as measured by a statistical comparison of their self-ratings in each language, but the IntHSs were more

strongly English-dominant. Lastly, the AdvHSs reported higher usage of and exposure to Spanish (with friends and parents) than the IntHSs.

In Section 4.3.3, I argued that the LCIs are an intermediary control group, and the between-group comparisons strongly support that designation. The LCIs' English self-ratings were higher than the SDCs' self-ratings and yet not significantly lower than the ratings of the HSs. With respect to English proficiency, therefore, the LCIs seem to rate themselves more proficient than the SDCs and slightly, though not significantly, less proficient than the HSs. If higher English proficiency predicts a higher likelihood of English to Spanish crosslinguistic influence, then the LCIs should be more susceptible to CLI than the SDCs but equally susceptible as compared to the HSs.

In Spanish, however, the LCIs pattern more closely with the SDCs in terms of both proficiency as well as language use and exposure. The LCIs did not differ from the SDCs in DELE scores, self-ratings in Spanish or reported proportions of use and exposure to Spanish. In comparison to the HS groups, the LCIs exhibited higher Spanish proficiency than the IntHSs (DELE and self-ratings) and the AdvHSs (self-ratings only) and, additionally, reported higher usage of and exposure to Spanish with friends as well as parents. The fact that the LCIs pattern with the HSs in English and the SDCs in Spanish suggests that they are the most balanced, in terms of proficiency, of the four bilingual groups. Strengthening this assertion is the fact that the LCIs demonstrated the smallest difference in Spanish and English SR's of the four bilingual groups in the study.

4.4 Procedure

All participants completed the study in the researcher's office, where they were greeted by the experimenter in English. After a short conversation (i.e., about the stresses

of the semester), always conducted in English, the experimenter presented participants with a consent form and then briefly explained each of the experimental tasks.

Following the completion of the consent forms, participants then completed the DELE proficiency exam, the Contextualized Elicited Production Task (CEPT), the Contextualized Acceptability Task (CAT), and the Mood Preference Task (MPT), always in that order. All experimental tasks included auditory (and also written stimuli) to reduce potential effects of HL literacy (Montrul, 2016a). Most participants completed the Language Background Questionnaire (LBQ) before completing the experiments. However, those participants who had not yet completed the LBQ at the time of the experiment completed it in the office (or at home) after finishing the MPT. When participants were completely finished with all experimental tasks, they received monetary compensation or extra credit for their time and effort. During the explanation of the consent form, participants were encouraged to stay afterwards if they had questions or comments about the study. Many participants, therefore, stayed and conversed for a few minutes afterwards.

4.4.1 DELE proficiency test

Participants completed a computerized version of the DELE on a large iMac computer in the office. Most participants completed the DELE in approximately 20 to 40 minutes. When they finished, the researcher would check their scores (i.e., whether they were classified as AdvHSs or IntHSs) to determine whether they would be assigned to complete Version A or Version B of the experiment. Version A and Version B tested the same variables but used different experimental items and item orders, as explained below.

The process of assigning participants a Version was not random (i.e., participants were assigned Version A or B based on how many participants in their respective

(proficiency) groups had completed each version thus far), but it did ensure that exactly half of the participants in each experimental group (except for the LCIs; n = 19) completed Version A of the experiment and the other half completed Version B.

4.4.2 Contextualized Elicited Production Task (CEPT)

Why test HSs' production of subjunctive mood morphology? As argued in a recent methodological paper by Ionin & Zyzik (2014), "the most direct information about language learners' linguistic knowledge is production" (p. 37). Of the many possible ways of testing production, it seems fairest to assess HSs' productive linguistic knowledge via oral, rather than written linguistic tasks (e.g., Montrul, 2016a). Given many HSs' limited Spanish literacy, unimodal written tasks may underestimate their knowledge of the HL. Oral production tasks, however, can provide HSs with more naturalistic, and therefore comfortable, contexts in which to exhibit their underlying grammatical knowledge.

The goal of the CEPT, which was completed using PowerPoint and an H4n Zoom recorder, was to test the participant groups' oral production of subjunctive and indicative mood forms as a function of (a) MoodSelectType (intensional or polarity), (b) ExpectedMood (subjunctive or indicative), and (c) PrimeStatus (baseline or prime). Further analyses also explored the role of VerbFrequency (frequent or infrequent). (Note that the names of predictive variables are capitalized throughout the remainder of the text.) To maximize the likelihood of eliciting HSs' subjunctive knowledge, the experiment was multimodal (e.g., auditory and visual), naturalistic, and highly contextualized.

In the CEPT, participants were asked to imagine themselves in a large department store where they are shopping for friends and family. For each experimental item, they listened, using headphones, to a brief context in Spanish which described a product (i.e., a

robot) with a given characteristic (i.e., with the ability to unload groceries from the car) that they were looking for within the store. After hearing each context, participants were presented with a sentence fragment, which they were asked to read out loud and complete orally to communicate their needs to a store clerk. To make sure participants understood the contexts, contexts were recorded by a bilingual, Spanish-dominant speaker (AofA Eng = 15 years) of Colombian Spanish, which is relatively unmarked (e.g., no –s elision). Participants had the option to replay the audio for all items. To ensure that memory did not significantly impact participants' performance, the text of both (a) the context and (b) the sentence fragment remained on screen after the presentation of the audio.

The items in the CEPT are divided into two Polarity and Intensional mood-selection categories. In the Polarity conditions, the presence or absence of presupposition in the experimental context is expected to guide participants' mood production.

When presupposition is clear, i.e., the participant finds the product that they are looking for, the target sentence fragment starts with '*Gracias, pero no necesito ayuda. Ya encontré un(a) X que…*' ('Thanks, but I do not need help. I already found a(n) X that…).

Due to the clear presupposition evident in such a context, participants in this condition are expected to produce indicative forms of the verb provided in parentheses. When there is no clear presupposition, i.e., the participant is not sure whether a product in question exists, the target sentence fragment starts with a lexical subject (i.e., *mi hermano*, 'my brother') followed by *quiere que le compre un(a) X que…*' ('_____ wants me to buy a(n) X that…'). In these cases, therefore, participants are expected to produce subjunctive mood forms of the verb provided in parentheses.

Table 4
Polarity mood selection: baseline conditions

+Presupposition ExpectedMood = Indicative	**-Presupposition** ExpectedMood = Subjunctive
CONTEXT "Necesitas un carro de control remoto con la capacidad de ir a más de 40 millas por hora. Es para tu nieto. Después de 5 minutos, *encuentras uno en la sección de juguetes*." ('You need a remote-control car with the ability to go over 40 mph. It's for your nephew. After five minutes, *you find one in the toy section*.'	CONTEXT "Necesitas un robot fuerte para bajar todas las compras del carro. Es para tu tía. *Sólo encuentras uno con la capacidad de cocinar*." ('You need a strong robot to unload groceries from the car. It's for your aunt. *You only find one with the ability to cook*.')
Le dices al dependiente: (You say to the clerk:	Le dices al dependiente: (You say to the clerk:
FRAGMENT "Gracias, pero no necesito ayuda. Ya encontré un carrito que (IR)…" ('Thanks, but I do not need help. I already found a car that (GO)…')	FRAGMENT "Mi tía quiere que le traiga un robot que (BAJAR)…" ('My aunt wants me to bring her a robot that (UNLOAD)…')
EXPECTED "… que **VA**$_{\mathbf{Indic}}$ a más de 40 millas por hora." ('…that GOES more than 40 mph.')	EXPECTED "…que **BAJE**$_{\mathbf{Subj}}$ las compras del carro." ('…that UNLOADS groceries from the car.')

In the – Presupposition condition, participants were exposed to both Baseline (n=6) and Prime (n=6) items. In the Baseline items (Table 4: - Presupposition), the sentence fragments included subjunctive mood forms (i.e., *mi tía quiere que le traiga*$_{SUBJ}$) which are lexically selected and therefore structurally distinct from adjectival relative clause subjunctive. In the Prime items (Table 5), on the other hand, the sentence fragments were coordinate structures with one verb in the subjunctive followed by the conjunction *y* ('and') and the complementizer *que*. The key difference between the Baseline and Prime conditions, therefore, was not whether participants were exposed to subjunctive mood forms (in both cases, they were exposed to subjunctive mood forms) but whether they were

exposed to a *structurally identical* form of subjunctive mood. Participants were expected to produce a higher proportion of subjunctive mood forms in the Prime condition.

Table 5.
Polarity mood selection: comparison of baseline and prime conditions

-Presupposition PrimeStatus = Baseline	**-Presupposition** PrimeStatus = Prime
CONTEXT "Necesitas un robot fuerte para bajar todas las compras del carro. Es para tu tía. Sólo encuentras uno con la capacidad de cocinar." ('You need a strong robot to unload groceries from the car. It's for your aunt. You only find one with the ability to cook.')	CONTEXT "Necesitas un robot fuerte *e inteligente* para bajar todas las compras del carro. Es para tu tía. Sólo encuentras uno con la capacidad de cocinar." ('You need a strong *and smart* robot to unload groceries from the car. It's for your aunt. You only find one with the ability to cook.')
Le dices al dependiente: (You say to the clerk:	Le dices al dependiente: (You say to the clerk:
FRAGMENT "Mi tía quiere que le compre un robot que (BAJAR)…" ('My aunt wants me to buy her a robot that (UNLOAD)…')	FRAGMENT "Busco un robot que sea inteligente y que (BAJAR)…" ('I'm looking for a robot that is smart and that (UNLOAD)…')
EXPECTED "…que **BAJE**$_{\mathbf{Subj}}$ las compras del carro." ('…that UNLOADS groceries from the car.')	EXPECTED "…que **BAJE**$_{\mathbf{Subj}}$ las compras del carro." ('…that UNLOADS groceries from the car.')

In the Intensional conditions, on the other hand, the presence or absence of presupposition does not affect the mood form that participants are expected to produce. (Consequently, there is no presupposition in any of the Intensional condition contexts.) What drives mood selection in the Intensional conditions is the complementizer which appears in the target sentence fragments. When the complementizer *porque* ('because') appears before the verb in parentheses, participants are expected to produce indicative mood forms. When the complementizer *para que* ('so that') appears, however, participants

are expected to produce subjunctive mood forms. For examples of contexts and sentence fragments of each type, see Table 6 below.

Table 6.
Intensional mood selection: indicative and subjunctive example items

PORQUE ExpectedMood = Indicative	**PARA QUE** ExpectedMood = Subjunctive
CONTEXT "Necesitas un juguete nuevo para tu gato, Pepe. Pepe es gordo y siempre pasa el día en la cocina mirando afuera, aburrido." ('You need a new toy for your cat, Pepe. Pepe is fat and always spends the day in the kitchen looking outside, bored.'	CONTEXT "Mi tío colecciona muchos libros. Necesita una escalera extensible para mover los libros de los estantes altos." ('My uncle collects lots of books. He needs an extendible ladder to move the books from the highest shelves.')
Le dices al dependiente: (You say to the clerk:	Le dices al dependiente: (You say to the clerk:
FRAGMENT "Busco un juguete nuevo porque mi gato siempre (MIRAR)…" ('I'm looking for a new toy because my cat always (LOOK)…')	FRAGMENT "Busco una escalera extensible para que mi tío (MOVER)…" ('I'm looking for an extendible ladder so that my uncle (MOVE)…')
EXPECTED "… porque mi gato siempre **MIRA**$_{\textbf{Indic}}$ afuera todo el día." ('…because my cat always LOOKS outside all day.')	EXPECTED "…para que mi tío **MUEVA**$_{\textbf{SUBJ}}$ los libros de los estantes más altos." ('…so that my uncle MOVES the books from the highest shelves.')

In the Para Que condition, participants were exposed to both Baseline (n=6) and Prime (n=6) items. In the Baseline items (Table 6: right column), the sentence fragments did not include any subjunctive mood forms. In the Prime items (Table 7), however, the sentence fragments appeared in a coordinate structure. That is to say, the fragments included the complementizer *para que* followed by a verb in subjunctive and then the target

verb in parentheses. Participants were expected to produce higher proportions of subjunctive mood morphology in the Prime condition than in the Baseline condition.

Table 7.
Intensional mood selection: comparison of baseline and prime conditions

PARA QUE PrimeStatus = Baseline	**PARA QUE** PrimeStatus = Prime
CONTEXT "Mi tío colecciona muchos libros. Necesita una escalera extensible para mover los libros de los estantes altos." ('My uncle collects lots of books. He needs an extendible ladder to move the books from the highest shelves.')	CONTEXT "Mi tío colecciona muchos libros. Necesita una escalera extensible para *alcanzar arriba* mover los libros de los estantes altos." ('My uncle collects lots of books. He needs an extendible ladder to *reach up* and move the books from the highest shelves.')
Le dices al dependiente: (You say to the clerk:	Le dices al dependiente: (You say to the clerk:
FRAGMENT "Busco una escalera extensible para que mi tío (MOVER)…" ('I'm looking for an extendible ladder so that my uncle (MOVE)…')	FRAGMENT "Busco una escalera extensible para que mi tío *alcance arriba* y (MOVER)…" ('I'm looking for an extendible ladder so that my uncle *reaches up* and (MOVE)…')
EXPECTED "…para que mi tío **MUEVA**$_{\mathbf{SUBJ}}$ los libros de los estantes más altos." ('…so that my uncle MOVES the books from the highest shelves.')	EXPECTED "…para que mi tío *alcance arriba* y **MUEVA**$_{\mathbf{SUBJ}}$ los libros de los estantes más altos." ('…so that my uncle reaches up and MOVES the books from the highest shelves.')

For each scenario in the CEPT (i.e., the robot that unloads groceries), the researcher created two separate items: a baseline version and a prime version. All participants saw either the baseline *or* the prime version for each scenario, but never both versions. When participants in Version A saw the baseline version of a given scenario, then participants in Version B saw the prime version (and vice versa).

For a summary of all experimental conditions in the CEPT, see Table 8 below.

Table 8.
Summary of experimental conditions in the CEPT

MoodSelection	ExpectedMood	Prime	Presupposition	Complementizer	Sample Fragment
Polarity (n=18)	Indicative (n=6)	Baseline (n=6)	+ Presupp	*QUE*	"Gracias, pero no necesito ayuda. Ya encontré un carrito que (IR)…"
	Subjunctive (n=12)	Baseline (n=6)	- Presupp	*QUE*	"Mi hermana quiere que le traiga una calculadora que (EXPLICAR)…"
		Prime (n=6)	- Presupp	*QUE*	"Busco una calculadora *que sea rosada* y que (EXPLICAR)…"
Intensional (n=18)	Indicative (n=6)	Baseline (n=6)	N/A	*PORQUE*	"Busco una tableta porque mi mamá siempre (LEER)…"
	Subjunctive (n=12)	Baseline (n=6)	N/A	*PARA QUE*	"Busco unos zapatos para que mi primo (HACER)…"
		Prime (n=6)	N/A	*PARA QUE*	"Busco unos zapatos *para que mi primo vaya afuera* y (HACER)…"

Figure 6 reveals the time course of an experimental item in the CEPT. First, the audio file plays, as the text appears on screen. Then, the prompt (*le dices al dependiente*, 'you say to the clerk') appears on screen followed by the sentence fragment and the target verb, always presented in parentheses. As mentioned above, participants' job was to record themselves finishing the fragment orally.

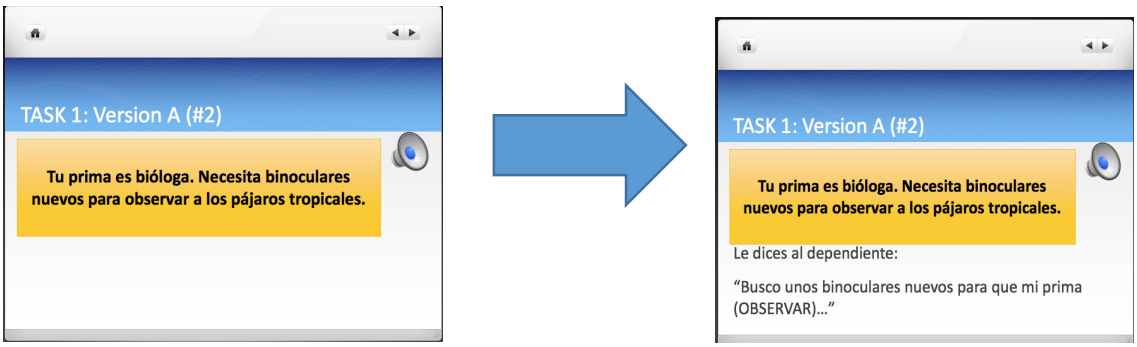

Figure 6: Example of experimental item from CEPT

The primary goals of the CEPT are to determine whether participants produce more subjunctive (a) with Intensional vs. Polarity mood selection and (b) in Prime as opposed to Baseline conditions. In order to make these comparisons as fair as possible, the researcher carefully selected target verbs according to verb frequency and verb regularity. Verb frequency rankings were taken from Davies' (2006) frequency dictionary, which is derived from a 20 million-word corpus of written (67%) and oral (33%) Spanish from both Spain (44%) and Latin America (56%), mostly between 1970 and 2000.

In both the Intensional condition and the Polarity condition, 8 of the 12 target verbs (66.7%) were regular while 4 of the 12 verbs (33.3%) were irregular. The verbs in the Intensional condition (Mean Frequency Rank: 267.42, SD: 180.32) were not significantly more frequent than the verbs in the Polarity condition (Mean Frequency Rank: 266.25, SD: 178.19), $t\ (46) = 0.23$, $p > .9$. Therefore, if participants perform better in the Intensional condition, as expected, then we can be confident that this improved performance is not attributable to verb frequency or verb regularity.

All participants saw the same 12 verbs in the Intensional condition, but those verbs were divided into two sets: Set 1 and Set 2. For participants completing Version A, Set 1

verbs appeared in the Prime condition and Set 2 verbs appeared in the Baseline condition. (The opposite was true for the participants completing Version B.) Of the six verbs in each set, four were regular and two were irregular. The verbs in Set 1 (Mean Frequency Rank: 271.33, SD: 183.95) were not significantly more frequent than the verbs in Set 2 (Mean Frequency Rank: 263.50, SD: 202.27), $t(10) = 0.07$, $p > .9$. Consequently, any differences between the Baseline and Prime conditions with intensional mood selection are not attributable to differences in verb frequency.

All participants also saw the same 12 verbs in the Polarity condition, again divided into two sets: Set 3 and Set 4. For participants completing Version A, Set 3 verbs appeared in the Prime condition and Set 4 verbs appeared in the Baseline condition. (The opposite was true for the participants completing Version B.) Due to a mistake by the experimenter, Set 3 contained one irregular verb and five regular verbs while Set 4 contained three irregular verbs and three regular verbs. Despite this incongruency, the verbs in Set 3 (Mean Frequency Rank: 270.67, SD = 192.95) were not significantly more frequent than the verbs in Set 4 (Mean Frequency Rank: 261.83, SD = 189.07), $t(10) = 0.08$, $p > .9$. Any differences between the Baseline and Prime conditions with polarity mood selection are not attributable, therefore, to differences in verb frequency.

In order to assess the effect of verb frequency on participants' subjunctive production, the researcher simply divided the 12 verbs in each subjunctive type (intensional and polarity) into the six most frequent (Frequent) and the six least frequent (Infrequent) verbs, according to Davies' (2006) rankings. In the statistical analysis, the researcher will assess the effect of frequency by comparing participants' subjunctive production with Frequent and Infrequent verbs, respectively.

To ensure that the Frequent and Infrequent verbs differed significantly in frequency ranking, independent samples t-tests were run for the 12 verbs appearing in the intensional subjunctive and polarity subjunctive conditions. With the intensional subjunctive verbs, frequency rankings for both Frequent and Infrequent verbs were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), although there was one outlier. There was homogeneity of variance, as assessed by Levene's test for equality of variances, $p = .566$. An independent samples t-test revealed that Frequent verbs, not surprisingly, had a significantly lower frequency ranking (M = 115.17, SD = 90.51) than the Infrequent verbs (M=419.67, SD = 104.71), $t(10) = 5.389$, $p < .001$.

With the polarity subjunctive verbs, frequency rankings for both Infrequent and Frequent verbs were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there were no outliers. There was homogeneity of variance, as assessed by Levene's test for equality of variances, $p = .854$. An independent samples t-test revealed that Frequent verbs, not surprisingly, had a significantly lower frequency ranking (M=122.17, SD = 106.28) than the Infrequent verbs (M=393.67, SD = 101.12), $t(10) = 4.533$, $p < .01$.

Prior to beginning the CEPT, participants completed two practice items (one targeting aspect, to ensure they would inflect the target verbs in parentheses; another with a coordinate PORQUE structure, to ensure that they understood what to do with the coordinate Prime condition items) with the experimenter to ensure that they fully understood the instructions (i.e., not adding modal verbs like *pueda*, 'is able to', before the verb in parentheses). Then, they were shown how to use the recorder so that they could record themselves producing each item in the task. When participants understood the instructions, they began the task, which took between 30-55 minutes.

4.4.3. Contextualized Acceptability Task (CAT)

Oral production tasks like the CEPT offer participants the chance to demonstrate their productive knowledge of subjunctive mood in a somewhat naturalistic way. Nonetheless, it would be unwise to assess HSs' knowledge of mood morphology via production tasks only, given (a) previous evidence of production/comprehension asymmetries in heritage bilinguals (e.g., Sherkina Lieber, 2015) and (b) the increased cognitive cost of producing, as opposed to comprehending, spoken language (e.g., Levelt, 1989). The Contextualized Acceptability Task, therefore, taps into HSs' *receptive* knowledge of subjunctive mood morphology by eliciting their judgments of highly contextualized examples of polarity indicative and subjunctive mood forms, respectively.

The goal of the CAT was to test whether participants' acceptance of subjunctive and indicative mood items was conditioned by the presence (or absence) of presupposition in the context. Therefore, the CAT focuses exclusively on Polarity mood contrasts.

As in the CEPT, participants in the CAT imagine themselves in a department store setting. This time though, instead of speaking, participants are asked to judge the acceptability of an animated character's Spanish productions. For each experimental item, participants watched and listened to interactions between Adriana and her young nephew, Nico, as they go holiday shopping. At the beginning of each interaction, Adriana appears on screen and tells Nico about a product on their shopping list, i.e., *una camisa* ('a shirt') with certain necessary characteristics (i.e., *con botones negros,* 'with black buttons') that she would like him to retrieve. After hearing her description, Nico appears on screen with a product retrieved from the store and asks Adriana, *Qué tal esta/o?* ('How about this one?'). Adriana then concludes the interaction by responding to Nico about the appropriateness of

the product he brings back. After watching the entire interaction carefully on screen via PowerPoint, participants rated the acceptability of Adriana's end-of-interaction responses, in context, using a 1 ('sounds very odd') to 5 ('sounds very good') scale.

In the FOUND scenario experimental items, which constitute exactly one half of the CAT, Nico brings back an appropriate product, thereby establishing presupposition in the context. In these items, Adriana responds to Nico by praising him (*Qué bien*, 'Great') and then restating the product that he found (i.e., *encontraste una camisa,* 'you found a shirt'), along with its pertinent characteristic (i.e., *que tiene*$_{INDIC}$ *botones negros*, 'that has black buttons'). Crucially, the pertinent characteristic is expressed by means of an adjectival relative clause headed by the complementizer, *que* ('that'). Due to the presupposition in FOUND scenario items, it is expected that participants will (a) <u>accept</u> Adriana's responses with <u>indicative</u> mood in the adjectival relative clause and (b) <u>reject</u> her responses with <u>subjunctive</u> mood forms in the adjectival relative clause.

In the WRONG scenario experimental items, which make up the other half of the CAT, Nico brings back an inappropriate product. By not bringing back an appropriate product, Nico ensures a lack of presupposition in the experimental context. Consequently, Adriana responds to Nico by answering negatively (*No, Nico*) and then restating the product of interest (*queremos una camisa*, 'we want a shirt') and its pertinent characteristic (*que tenga*$_{SUBJ}$ *botones negros*, 'that has black buttons'). As with the FOUND scenario items, the pertinent characteristic is expressed with an adjectival relative clause headed by *que*. Due to the lack of presupposition in WRONG scenario items, participants are expected to (a) <u>accept</u> Adriana's responses with <u>subjunctive</u> mood in the adjectival relative clause and (b) <u>reject</u> her responses with <u>indicative</u> mood forms in the adjectival relative clause.

In summary, the two within-group independent variables Scenario (Found, Wrong) and Mood (Indicative, Subjunctive) combine to create a total of four experimental conditions: FoundIndicative, *FoundSubjunctive, *WrongIndicative and WrongSubjunctive. (Infelicitous conditions are marked with asterisks.) A summary of the four conditions is presented in Table 9.

Table 9.
Summary of experimental conditions in the CAT

Scenario	Relative Clause Mood	# of Items	Expected Response
FOUND (+Presupposition)	Indicative	6	Accept
	*Subjunctive	6	Reject
WRONG (-Presupposition)	*Indicative	6	Reject
	Subjunctive	6	Accept

If participants are sensitive to the relationship between mood morphology and presupposition, we should see a two-way interaction in their acceptance patterns whereby they show higher acceptance of indicative mood items in the FOUND scenario and higher acceptance of subjunctive mood items in the WRONG scenario. If, on the other hand, participants show a mood bias (i.e., default preference for indicative or subjunctive forms), we might expect them to show higher acceptance of a single mood in both conditions.

Figure 7 presents the time course of an experimental item, in this case, from the FoundIndicative condition. Screen shots are presented to show the progression of the slide.

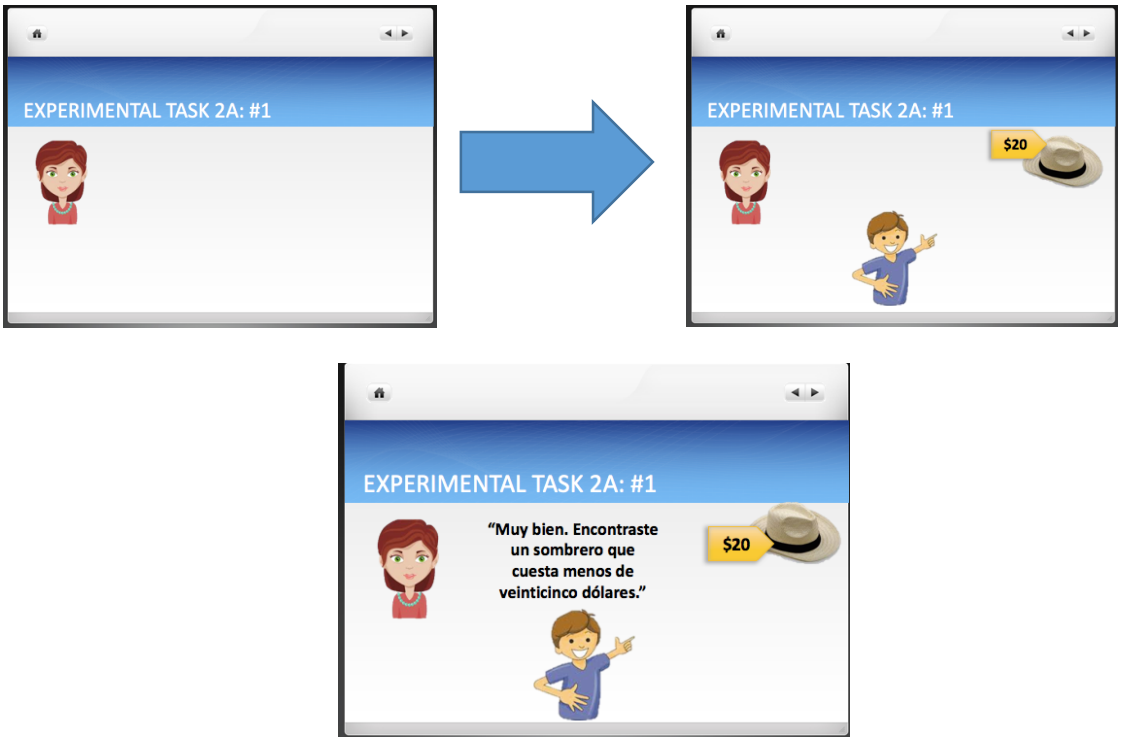

Figure 7: Example of experimental item from CAT

First, Adriana appears in the top left. As she appears, she tells Nico: "*Necesitamos un sombrero por menos de $25*." ('We need a hat for less than $25.') Then, Nico appears on screen, along with the product that he found, in this case, a hat for $20. As he appears on screen, he asks Adriana, "¿Qué tal este?" ('How about this one?). Finally, Adriana reappears on screen and responds by clarifying what Nico should be bringing back or, in this case, praising him for having brought back an appropriate item. The manipulation, as mentioned above, is the mood form that Adriana produces in the adjectival relative clause.

There were a total of 24 experimental items in the CAT. In the 24 experimental items, only six verbs were used, each four times: *ser* ('be'), *estar* ('be'), *tener* ('have'), *venir* ('come'), *costar* ('cost') and *funcionar* ('work'). These six verbs were carefully selected because they are relatively natural within the shopping context of the CAT. Presenting all participants with each of the six verbs in each of the four conditions has one key advantage:

this design allows me to evaluate whether participants are more sensitive to mood distinctions with certain verbs. It is quite possible, for example, that participants might be more sensitive to mood distinctions with more frequent verb forms. (That is to say, participants may be more sensitive to the *tiene*/*tenga* than the *cuesta/cueste* distinction.)

Table 10.
Summary of filler conditions in the CAT

Scenario	Main Clause Aspect	# of Items	Expected Response
FOUND (+Presupposition)	+ Perfective (*Encontraste*)	4	Accept
	- Perfective (**Encontrabas*)	4	Reject
WRONG (-Presupposition)	+Perfective (**Quisimos*)	4	Reject
	- Perfective (*Queríamos*)	4	Accept

To conclude the description of the CAT, I now introduce a few crucial methodological details about the fillers, the two versions of the task, the practice items, and the voicing of Adriana.

In addition to the 24 experimental items, there were 16 fillers (Table 10) which targeted the acceptability of aspectual morphology in the matrix clause of Adriana's responses. In the FOUND scenario fillers, Adriana always employs the telic verb, *encontrar* ('find') in the matrix clause of her response. Given that Nico has found the right product in these items, the perfective form *encontraste* ('you found') is acceptable here while the imperfective form *encontrabas* ('you were finding') is not. In the WRONG scenario fillers, on the other hand, Adriana always employed the stative verb, *querer* ('want') in the matrix clause of her response. Since Nico has *not* found the right product in these experimental items, it is now the imperfective form *queríamos* ('we wanted') which

is acceptable and the perfective form *quisimos* ('we wanted$_{+PERF}$') which is not acceptable. The importance of these fillers in particular is that they direct participants' attention (consciously or unconsciously) towards the verbal morphology in Adriana's responses.

There were two versions of the CAT: Version A and Version B. Both versions had the same order of items (i.e., #1: FoundIndicative with *costar*) and the same product interactions (i.e., Nico correctly bringing back a hat for less than $25). Where they differed was in the mood forms that appeared with each of these interactions. If a mood form in a given interaction appeared in the indicative in Version A, it appeared in the subjunctive in Version B and vice versa. By designing the experiment in this way, the researcher is able to ensure that differences in acceptability ratings are determined by mood morphology rather than by other irrelevant elements of each scenario.

Before each participant completed the CAT, the researcher carefully explained the instructions. First, I introduced participants to Nico and Adriana. Then, I explained that Nico was helping Adriana by listening to her describe products on her shopping list and then bringing them back from within the store. I clarified that after Nico comes back, Adriana either praises him for finding the right product or clarifies the type of product that he needs to bring back. Finally, I told them that their job was to listen to Adriana's response and rate how good it sounds in that context.

To test whether participants understood the instructions, I stood nearby as they completed two practice items at the beginning of their CAT answer sheet, both of which targeted aspectual mood morphology. The first practice item was a WRONG scenario filler with target, imperfective aspect. The second practice item was a FOUND scenario filler with non-target, imperfective aspect. Most participants correctly accepted the first item and

rejected the second item. After participants completed both practice items, I asked them whether they understood the task. If they did not, I explained it again until they understood the instructions. Finally, I presented a reminder slide on the PowerPoint which reminded participants to (a) watch the interactions carefully from start to finish and (b) pay attention to Adriana's *entire response*. (This reminder was added to make sure that participants did not just make rapid acceptability judgments based on the form of the matrix verbs *encontrar* and *querer*). Participants then put on headphones and began the CAT.

Adriana's character was voiced by an L1 Spanish-L2 English late bilingual from Mexico with high proficiency in English. All of her "lines" for Version A and Version B were recorded in an office and then rerecorded as necessary to fix any acoustic problems (i.e., fuzziness in the sound file). The speaker read Adriana's lines with a lot of enthusiasm and vigor in order to make the CAT (which can be very repetitive to complete) a bit more engaging and entertaining for participants.

4.4.4. Mood Preference Task (MPT)

Even when highly contextualized, acceptability judgment tasks like the CAT also run the risk of underestimating HSs' sensitivity to subtle grammatical distinctions (Polinsky, 2006) such as distinctions between subjunctive and indicative mood morphology. In general, HSs demonstrate a robust tendency to "overaccept" ungrammatical/infelicitous items (Polinsky, 2016). This pattern likely makes HSs appear less sensitive to certain grammatical properties than they actually are. An additional challenge of using acceptability tasks with HSs (and other populations, too) is that researchers cannot be sure that participants are accepting or rejecting sentences in direct response to the relevant experimental manipulations (Ionin & Zyzik, 2014).

In the Mood Preference Task (MPT), therefore, the researcher seeks to overcome these potential problems by presenting participants with *pairs* of sentences, still highly contextualized, which differ only in their mood morphology. By doing so, the researcher hopes to both (a) mitigate the potentially impactful role of HSs' acceptability bias and (b) isolate the structure of interest, increasing the likelihood that participants' responses are driven by mood morphology, rather than other potentially confounding factors.

The goal of the Mood Preference Task (MPT) was to evaluate participants' mood preferences with both intensional and polarity mood selection. As such, the MPT tests whether participants, when presented with both mood options, prefer indicative or subjunctive (a) with *para que* ('so that') and *porque* ('because') as well as (b) with and without contextual presupposition. Like previous tasks, the MPT takes place within a department store setting. This time, the two characters of interest are Gabriela and Elena, identical twins who are on a shopping spree for friends and family members. For each experimental item, participants hear a context sentence in which Gabriela presents introductory information about a product that she and Elena have either already bought or are still looking for. After the context sentence, both sisters provide follow-up sentences. Participants' job in the MPT is to select the follow-up sentence that sounds best to them.

As in the CEPT, experimental items in the MPT are divided into two major categories on the basis of the type of mood selection involved. In Polarity items, target mood selection is determined by the presence or absence of presupposition in the context. In Intensional items, on the other hand, target mood selection is determined by the lexical selection properties of the complementizer. In both cases, the two sentences which

participants hear for each experimental item constitute minimal pairs, differing only in the mood morphology of the target verb.

There are two types of Polarity items: + Presupposition and – Presupposition. In the + Presupposition items, the context sentence provides a few key pieces of information. First, Gabriela reveals that she and her sister *needed* to buy something for a particular person, i.e., their father. Then, she describes a place where they *went*, i.e., the sporting goods store. Together, the past tense morphology on the verbs in each of these sentences implies that they have already found a particular product for the given person and therefore, are no longer looking for something else.

Table 11.
Example item from polarity, +presupposition condition

POLARITY: + Presupposition Condition	
CONTEXT SENTENCE *"Necesitábamos algo para nuestro papá. Fuimos a la tienda de deportes y*: ('We needed something for our dad. We went to the sporting goods store and:	
GABRIELA's FOLLOW-UP *"Encontramos una maleta que* ***lleva***$_{INDIC}$ *muchísimos balones de fútbol."* ('We found a bag that carries lots of soccer balls.') **PREFERRED**	ELENA's FOLLOW UP *"Encontramos una maleta que* ***lleve***$_{SUBJ}$ *muchísimos balones de fútbol."* ('We found a bag that carries lots of soccer balls') NOT PREFERRED

Their follow-up sentences in the + Presupposition condition (Table 11) begin with the telic verb *encontramos* ('we found'), followed by the product they found (i.e., *una maleta*, 'a bag') and an adjectival relative clause headed by *que*. The only difference between their follow-up sentences is the mood morphology of the verb in the adjectival relative clause. In this condition, we expect that participants will prefer the indicative mood form given the clear presupposition in the context.

In the – Presupposition condition (Table 12), the context sentences do *not* imply presupposition. In the first context sentence of the – Presupposition items, Gabriela states that she and her sister need something (i.e., *algo dulce*, 'something sweet') for a particular person. Crucially, she uses the present tense form of the verb *necesitar*, implying that at the time of speaking, they are still in need of a suitable product for that person. In the next context sentence, she reinforces this fact by stating, *no hemos visto nada* ('we have not seen anything'). Up to this point, it is quite clear that the sisters have not found a suitable product and therefore are still searching for one.

Table 12.
Example item from polarity, - presupposition condition

POLARITY: - Presupposition Condition	
CONTEXT SENTENCE *"Necesitamos algo dulce para nuestro papa pero no hemos visto nada*: ('We need something sweet for our dad, but we have not seen anything:	
GABRIELA's FOLLOW-UP *"Buscamos un postre que* ***tiene***$_{INDIC}$ *chocolate y arándanos."* ('We're looking for a dessert that has chocolate and blueberries') NOT PREFERRED	ELENA's FOLLOW UP *"Buscamos un postre que* ***tenga***$_{SUBJ}$ *chocolate y arándanos."* ('We're looking for a dessert that has chocolate and blueberries') **PREFERRED**

The follow up sentences in the – Presupposition condition, therefore, always begin with the verb *buscamos* ('we are looking for'), followed by the product of interest (i.e., *un postre*, 'a dessert') and an adjectival relative clause headed by *que.* The only difference between the two follow-up sentences is the mood morphology on the verb in the adjectival relative clause. In this condition, participants should prefer subjunctive mood morphology in the relative clause, given the clear lack of presupposition in the experimental context.

In the Intensional conditions, mood selection on the target verb is determined by the selecting complementizer and crucially *not* by any information in the preceding context

sentences. Therefore, the context sentences for both the PARA QUE (Table 13) and PORQUE (Table 14) conditions share the same format. In all of the Intensional context sentences, Gabriela uses a command (*mira*, 'look') to direct attention to a particular product that the twins have bought for someone, i.e., *la bicicleta que le compramos a nuestro papá* ('the bike that we bought for our dad').

In the PARA QUE condition, the follow-up sentences always include an adverbial clause headed by the complementizer, *para que* ('so that'). The only difference between the two follow-up sentences, therefore, is the mood morphology on the verb following *para que*. Because *para que* obligatorily selects for subjunctive mood morphology, we would expect participants in this condition to prefer follow-up sentences with subjunctive mood.

Table 13.
Example item from intensional para que *condition*

INTENSIONAL: *Para Que* Condition	
CONTEXT SENTENCE *"¡Mira la bicicleta que le compramos a nuestro papa*!" ('Look at the bike that we bought for our dad!')	
GABRIELA's FOLLOW-UP *"Se la compramos para que él* ***haga***$_{SUBJ}$ *más ejercicio."* ('We bought it for him so that he exercises more') **PREFERRED**	ELENA's FOLLOW UP *"Se la compramos para que él* ***hace***$_{INDIC}$ *más ejercicio."* ('We bought it for him so that he exercises more') NOT PREFERRED

In the PORQUE condition, the follow-up sentences always include an adverbial clause headed by the complementizer, *porque* ('because'). The only difference between the two follow-up sentences for each item, therefore, is the mood morphology on the verb following *porque*. Because *porque* selects for indicative mood morphology, we would expect participants in this condition to prefer follow-up sentences with indicative mood.

Table 14.
Example item from intensional porque *condition*

INTENSIONAL: *Porque* Condition	
CONTEXT SENTENCE *"Mira el cuaderno grande que le compramos a nuestra mamá!"* ('Look at the big notebook we bought for our mom!')	
GABRIELA's FOLLOW-UP *"Se lo compramos porque ella* ***necesita***$_{INDIC}$ *cuadernos para su poesía."* ('We bought it for her because she needs notebooks for her poetry') **PREFERRED**	ELENA's FOLLOW UP *"Se lo compramos porque ella* ***necesite***$_{SUBJ}$ *cuadernos para su poesía."* ('We bought it for her because she needs notebooks for her poetry') NOT PREFERRED

A summary of the experimental conditions is provided below in Table 15.

Table 15.
Summary of experimental conditions in MPT

MoodSelectionType	MoodSelector	# of Items	Preferred Mood
POLARITY	+ Presupposition	5	Indicative
	- Presupposition	5	Subjunctive
INTENSIONAL	*Porque*	5	Indicative
	Para Que	5	Subjunctive

Figure 8 is an example of an experimental item from the MPT, presented to help visualize the participants' experience completing this task. First, participants hear the context sentence, which is also presented in a text box in the center of the screen. After a two-second pause, Gabriela provides her follow-up sentence. As she speaks, the gold "name plate" under her picture expands to ensure that participants know that she is speaking. Following another two-second pause, Elena reads her follow-up sentence. Her gold name plate also expands to show she is talking.

After listening to both follow-up sentences, participants respond on their answer sheets by circling the picture of the sister whose sentence sounded better. The pictures of

Gabriela and Elena on the answer sheet are exactly the same (and in the same order) as in the PowerPoint slides, making the task of identifying the characters very easy.

Figure 8: Example of experimental item from MPT

To conclude my discussion of the MPT, I now present a few methodological considerations taken into account during the experimental design. With regards to verb selection, all twenty verbs in the MPT were selected from the verbs which were used in the CEPT. (All verbs that appeared in both tasks always appeared in the same condition, i.e., Intensional/SubjunctiveExpected or Polarity/IndicativeExpected) This was done in order to facilitate follow-up comparisons of participant accuracy with the same verbs across different task modalities. (For example, it may be the case that participants are significantly

more accurate preferring *para que corra*$_{SUBJ}$ ('so that he runs') than they are at producing it.) We will see the benefits of this decision in the Chapter 6.

Table 16.
Summary of filler conditions in the MPT

Filler Type	# of Items	Competing Forms
Psychological Predicates (*gustar* with plural themes)	5	i. **gusta* ii. *gustan*
Prepositional Particles (*buscar* with(out) particles)	5	i. *buscamos* ii. **buscamos por*

In addition to the twenty experimental items, there were also ten filler sentences (Table 16), which targeted verbal agreement (with the psychological predicate, *gustar*, 'to be pleasing to') and argument structure (with the verb *buscar*, 'to look for'). The sole purpose of these experimental items was to distract participants from paying explicit attention to mood morphology in Spanish.

There were two versions of the MPT: Version A and Version B. The only difference between the versions was the order of the experimental items. The order of items in Version A was the exact opposite of the order of the items in Version B. By having inverse item orders in Version A and Version B, the experimenter hoped to cancel out any potential adverse effects of participant boredom and/or fatigue on accuracy in the MPT, which was always the final task completed by participants. (If tired participants stop paying careful attention for the last 15 items, then this will affect all items equally since the last 15 items in Version A are the first items in Version B.)

Finally, the voices for both Gabriela and Elena were provided by an L1-Spanish/L2-English late bilingual (AofA Eng= 13 years) with high English proficiency. The same speaker voiced both characters so that participants' mood preferences on any given item would not be attributable to biases against acoustic characteristics of the speaker's voice.

(If two speakers had been used, participants may have shown an overall bias to prefer items read by Speaker 1 vs. Speaker 2.) Of course, using the same voice for both characters also made sense within the context of the MPT, given that the two main characters were supposed to be identical twin sisters.

4.5 Chapter summary

In this chapter, I began by introducing the broader goals of the dissertation, namely, to describe HSs' knowledge of intensional and polarity subjunctive mood morphology and identify between- and within-group factors which shape that knowledge. In Section 4.2, I presented the specific RQs guiding the present study as well as theoretically and empirically motivated hypotheses for each. In Section 4.3, I introduced the four groups of bilingual participants in the present study (Table 17 below), providing information about their Spanish proficiency, age of acquisition of English, and self-reported language usage in both English and Spanish.

Table 17.
Summary of participant groups

Participant Group	Age at Time of Testing	AofA English	DELE Proficiency	Spanish Self-Rating	English Self-Rating
AdvHSs (n=22)	Mean: 21.22 SD: 2.60	Age 6 or earlier	Mean: 42.41 SD: 1.74	Mean: 8.41 SD: 1.37	Mean: 9.82 SD: 0.66
IntHSs (n=20)	Mean: 19.80 SD: 1.06	Age 6 or earlier	Mean: 34.80 SD: 2.69	Mean: 6.75 SD: 1.77	Mean: 9.55 SD: 1.19
LCIs (n=19)	Mean: 20.73 SD: 2.77	Between Age 8 and Age 12	Mean: 44.00 SD: 3.61	Mean: 9.58 SD: 0.77	Mean: 8.89 SD: 1.15
SDCs (n=20)	Mean: 30.37 SD: 7.25	Age 13 or later	Mean: 45.60 SD: 2.62	Mean: 9.95 SD: 0.22	Mean: 7.70 SD: 1.34

Finally, in Section 4.4, I presented the methodology, beginning with the larger goals of each task before highlighting crucial elements of each task's experimental design. The Contextualized Elicited Production Task (CEPT) was designed to determine whether

participants' oral production of subjunctive mood morphology was significantly impacted by the within-group factors of structural priming, lexical selection, and lexical frequency, respectively. The Contextual Acceptability Task (CAT) complemented the CEPT by providing participants with the opportunity to demonstrate *receptive* knowledge of indicative and polarity mood forms, respectively. Lastly, the Mood Preference Task (MPT) offered a less taxing, and more direct, evaluation of participants' receptive mood sensitivity. Throughout Section 4.4, I argue that the use of a more versatile, multimodal array of experimental tasks offers HSs the best chance to reveal their knowledge, however inconsistent, of subjunctive mood forms in Spanish.

The four experimental groups were carefully selected to address the critical role of between-group variables on knowledge of subjunctive mood. By including participants with early (HSs), middle (LCIs) and late (SDCs) ages of acquisition of English, I am able to evaluate the extent to which participants with earlier exposure to English are more vulnerable to variability in their knowledge of subjunctive. By testing HSs with both advanced and intermediate proficiency, I also evaluate the role of proficiency in HSs' variable knowledge of subjunctive mood.

In Chapter 5, I present the results from the three experimental tasks. In each section, I will present the raw results, followed by the results from a series of Generalized Linear Mixed Models (GLMMs). Further discussion of the results will be presented in Chapter 6.

CHAPTER 5:
RESULTS

5.1 Introduction

In this chapter, I will present and analyze the results of the Contextualized Elicited Production Task (CEPT; Section 5.2), the Contextualized Acceptability Task (CAT; Section 5.3), and the Mood Preference Task (MPT; Section 5.4). In doing so, I will shed light on the three RQs presented in Section 4.2 and briefly summarized here.

First, what is the nature of HSs' productive and receptive knowledge of subjunctive mood morphology? Second, how do the *between-group* factors (i) age of acquisition of English, (ii) Spanish proficiency and (iii) use of Spanish impact HSs' knowledge of subjunctive? Lastly, how do the *within-group* factors (i) age of acquisition of the property, (ii) priming status and (iii) verb frequency affect HSs' knowledge of subjunctive?

5.2 Contextualized Elicited Production Task (CEPT)

Participants in the CEPT listened to contextualized scenarios and then finished (orally) incomplete sentences which targeted subjunctive and indicative mood forms, respectively. The results of the CEPT help to answer the RQs. By eliciting participants' knowledge of mood and analyzing the internal systematicity of that knowledge, the CEPT provides a descriptive picture of HSs' productive mood knowledge (RQ #1). By testing the roles of (a) AofA Eng and (b) proficiency on mood production, the CEPT addresses the role of between-group factors on HSs' expressive mood knowledge (RQ #2). Finally, by examining the effects of (a) mood selection type, (b) priming status, and (c) lexical frequency, the CEPT evaluates the role of three within-group factors on HSs' mood knowledge (RQ #3).

In the sub-sections that follow, I will present the descriptive statistics (5.2.1), the statistical models used (5.2.2) and, finally, the results from the statistical analyses (5.2.3).

5.2.1 Descriptive statistics

In this section, I will briefly introduce the variables in this experiment. Immediately thereafter, I will discuss the way in which the experimental data was trimmed, explaining and justifying the exclusion of certain participant responses. Using the trimmed data, I will achieve two goals: first, I will argue that the design of the experiment works, and second, I will highlight a characteristic of the data which will be important in Section 5.2.3.

5.2.1.1 Variables in the CEPT

The dependent variable in the CEPT is SubjunctiveProduction, that is to say, the proportion of the time that participants produce subjunctive mood forms. Because participants either produce subjunctive or indicative, this variable is binary, which will be critical in the selection of appropriate statistical models.

The independent variables in the study can be subdivided into between-group and within-group variables. The between-group factors include Group (Advanced Heritage Speakers (AdvHSs), Intermediate Heritage Speakers (IntHSs), Late Childhood Immigrants (LCIs) and Spanish-dominant Controls (SDCs)), which includes both AofA Eng and Spanish proficiency. The within-group factors include MoodSelectType (Intensional or Polarity), PrimeStatus (Prime or Baseline), ExpectedMood (Subjunctive or Indicative), and Verb Frequency. In those statistical analyses which include only those items where subjunctive mood morphology is expected, (Section 5.2.3), VerbFrequency (Frequent or Infrequent) will also be a within-group independent variable.

5.2.1.2. Data trimming and exclusions

All 81 participants completed the CEPT, which included 36 experimental items and 18 fillers. However, because this task required participants' careful attention, not all participants successfully completed all experimental items in the CEPT. Of the 2,916 total experimental responses, 160 (5.5%) were removed from the analysis for the ten reasons presented below in Table 18. In general, the reasons for excluding a participant's response can be divided into two categories. First, participants' responses were excluded when the researcher, due to technical errors, ambiguous pronunciation, or ambiguous morphology, could not confidently label a response as indicative or subjunctive mood. Second, participants' responses were excluded when they problematized the statistical analysis, either by adding new lexical material, e.g., modal verbs before the target verb or different target verbs altogether or inflecting the target verb in a non-present form. Responses with new lexical material have to be removed in order to run Item as a random factor in the statistical analysis. Responses with non-present inflections of the target verb have to be removed in order for the dependent variable MoodProduction to remain binary.

The excluded items were produced by participants from all four of the experimental groups including the AdvHSs (12 excluded items produced by 8 of 22 AdvHSs), IntHSs (51 excluded items produced by 15 of 20 IntHSs), LCIs (51 excluded items produced by 16 of 19 LCIs) and SDCs (46 excluded items produced by 14 of 20 SDCs).

Table 18.
Items excluded from CEPT analysis

Reason for Exclusion	Example From Experiment	# of Exclusions
1. Modal Verb: Participants added modal verb prior to the target verb in parentheses.	"Busco un robot que sea verde y que *puede* hablar en cuatro idiomas." Participant #55, Item #32: IntHS	n=39
2. Recording Errors: Due to technical error, experimental item was not saved.	N/A	n=38
3. Wrong Lexical Item: Participants did not use the target verb form in answer.	"Ya encontré una nevera que *diga* (target verb: decidir) si la comida todavía está buena o no." Participant #75, Item #44: LCI	n=15
4. Non-Present Inflection: Participants used non-present form of target verb.	"Busco un robot que sea inteligente y que *bajará*$_{FUT}$ las compras del carro." Participant #74, Item #8: AdvHS	n=10
5. Non-Finite Verb Form: Participants used infinitive form of the target verb.	"Busco una nevera que cueste menos de $500 y que *elegir* comidas para la cena." Participant #52, Item #25: IntHS	n=10
6. Inserted Lexical Items Prior to Target Verb: Participants added lexical items before target verb.	"Busco un juguete nuevo porque mi gato *se lo pasa* mirando afuera aburrido." Participant #55, Item #26: IntHS	n=7
7. First Person Inflection: Participants interpreted the relative clause such that the subject was first person.	"Ya encontré un reloj que *necesito*." Participant #58, Item #29: AdvHS	n=6
8. Ambiguous Vowel Pronunciation: The researcher could not determine if participants produced [a] or [e] at end of the target verb inflection.	"Ya encontré un estuche que *llev[a/e]* mis dos guitarras eléctricas." Participant #86, Item #5: SDC	n=4

9. Ambiguous Mood Morphology: Participants used a subjunctive root (e.g., *hag*-) but indicative inflection (e.g., [e] for –er.)	"Busco unos zapatos para que mi primo *hague* más ejercicio." Participant #18, Item #43: IntHS	n=2
10. More Than One/Other: Participants' response was excluded (a) for more than one of the above reasons, or (b) for another reason.	(a) "Ya encontré una aspiradora que *seguía buscando*." (Non-present + 1st per) Participant #69, Item #5: SDC (b) "Busco el despertador que cuesta menos de $75 y que produce el sonido de las olas." (Definite determiner used instead of indefinite determiner, *un.*) Participant #1, Item #52: IntHS	n=29
		Total: 160

5.2.1.3. Descriptive results

After excluding these 160 items, 2756 items remained for statistical analysis. I will briefly discuss these data to make two important points. First, the SDCs perform as expected in all experimental conditions, validating the design of the experiment. Second, the SDCs perform invariably (e.g., 0% or 100%) in multiple experimental conditions, a fact which will have important consequences for the modeling of the data in Section 5.2.3.

As demonstrated in Table 19, the SDCs perform as expected in the Intensional Mood selection conditions. In the Intensional Subjunctive conditions, where the complementizer *para que* triggers subjunctive mood morphology, the SDCs produced subjunctive 100% of the time. In the Intensional Indicative condition, where the complementizer *porque* selects for indicative mood morphology, the SDCs did not produce a single instance of subjunctive mood morphology. Together, the results in these conditions

demonstrate that the SDCs have categorical and invariant knowledge of the mood selectional requirements of the complementizers *para que* and *porque*.

Table 19.
Participants' use of subjunctive by Group, Condition

Participant Group	Mood Select Type	Expected Mood	Prime Status	Mood Production		% of Subjunctive Production
				Subjunctive (Total #)	Indicative (Total #)	
SDC	Intensional	Indicative	Baseline	0	111	0.0%
		Subjunctive	Baseline	110	0	100.0%
			Prime	116	0	100.0%
	Polarity	Indicative	Baseline	16	91	15.0%
		Subjunctive	Baseline	110	3	97.3%
			Prime	117	0	100.0%
LCI	Intensional	Indicative	Baseline	0	110	0.0%
		Subjunctive	Baseline	105	0	100.0%
			Prime	108	1	99.1%
	Polarity	Indicative	Baseline	42	51	45.2%
		Subjunctive	Baseline	108	2	98.2%
			Prime	110	0	100.0%
AdvHS	Intensional	Indicative	Baseline	1	131	0.8%
		Subjunctive	Baseline	104	27	79.4%
			Prime	101	31	76.5%
	Polarity	Indicative	Baseline	25	101	19.8%
		Subjunctive	Baseline	89	41	68.5%
			Prime	88	41	68.2%
IntHS	Intensional	Indicative	Baseline	6	110	5.1%
		Subjunctive	Baseline	53	52	50.5%
			Prime	52	55	48.6%
	Polarity	Indicative	Baseline	16	90	15.1%
		Subjunctive	Baseline	34	84	28.8%
			Prime	37	75	33.0%
				1548	1208	56.2%

In the Polarity Mood selection conditions, the SDCs also perform as expected, if also somewhat more variably. In the Polarity Subjunctive conditions, where the lack of

contextual presupposition should lead to subjunctive production, the SDCs produced subjunctive 98.7% of the time. In the Polarity Indicative condition, where the presence of contextual presupposition should lead to indicative production, the SDCs only produced subjunctive 15% of the time. The results from the two Polarity Mood selection conditions reveal two main trends. First, the SDCs are clearly sensitive to contextual presupposition when producing (or not producing) subjunctive mood morphology. Secondly, the SDCs' mood sensitivity is slightly more variable with Polarity than with Intensional mood selection. This variability is not surprising, however, given that only the Polarity items require participants to pay attention to the experimental contexts in order to produce target subjunctive or indicative mood morphology, respectively.

Before preceding to the statistical models, it is important to point out a critical characteristic of the SDCs' raw data. In four of the six experimental conditions, the SDCs' data is totally invariant, meaning that all SDCs produced the same mood form for every item of those conditions. While this invariant data clearly illustrates the consistency of the SDCs' mood production, it has consequences for the statistical analysis. In order to run Generalized Linear Mixed Models (GLMMs) on the data set, it will be necessary to alter a few data points in each of the invariant conditions, as discussed below.

5.2.2 Statistical models

The data from the CEPT are analyzed using Generalized Linear Mixed Models (henceforth, GLMMs). In Section 5.2.2.1, I present a brief introduction to GLMMs and their appropriateness for experiments with binary dependent variables.

5.2.2.1 Introduction to GLMMs

In the present study, data are analyzed GLMMs. GLMMs, like ANOVAs, allow researchers to evaluate the statistical impact of between- and within-group variables on participants' performance with the dependent variable. Unlike ANOVAs, however, GLMMs also allow researchers to control for random variance both among subjects (i.e., some participants produce more subjunctive) and lexical items (i.e., some verbs/experimental items elicit more subjunctive from participants than other items)[25].

It is important to briefly outline what GLMMs generate during statistical analysis. GLMMs generate the *predicted probability* of different binary outcomes, in the case of the CEPT, whether a given experimental item is produced with or without subjunctive mood morphology. The probabilities are expressed in log-odds, which can be easily converted into odds (e.g., "2 to 1" or, numerically, 2.0) or traditional probabilities (e.g., (2/ (2+1) = 66.7%). When comparing two participant groups' production of subjunctive in a given condition, the GLMM will estimate the odds that participants in each group will produce subjunctive with items in that condition. An F-test on the difference in odds between the groups will produce a *p*-value, which clarifies (a) whether there are statistically significant differences in the groups' odds of producing subjunctive in that condition, or (b) whether there are differences in one group's predicted odds across different conditions.

It is important to note, however, that *p*-values do not provide information about the *size* of the group differences, e.g., *how much* the groups' odds differ. For that, we must calculate odds-ratios (OR's), which are transparent, non-standardized measures of effect size (Durlak, 2009). For example, if AdvHSs' odds of producing subjunctive in a Condition

[25] For more on the advantages of GLMMs, as well as the importance of not using ANOVAs with binary dependent variables, see Jaeger (2008) and sources within.

X are 4 to 1 (80% probability) and the SDCs' odds are 19 to 1 (95%), we can calculate the effect size (OR) of the difference between the two groups by dividing the SDCs' odds (19) by the AdvHSs' odds (4) to get an OR of 4.75. This figure tells us that the SDCs' odds of producing subjunctive are nearly five times greater than the AdvHSs' odds. OR's, therefore, reveal the *size* of the difference between groups. OR's that are closer to 1 indicate greater similarity between the two groups/conditions being compared while OR's farther from 1 indicate lesser similarity.

5.2.3. Results from the GLMMs

In sections 5.2.3.1 through 5.2.3.3, I will present the results from three GLMMs used to analyze the CEPT data. The title of each sub-section will indicate the combination of independent variables that are examined within a particular statistical analysis. In each section, I will start by exploring both main effects and interaction effects (both two-way and three-way). I will present the results in tables which include log-odds, standard errors, and odds, as well as predicted probabilities and 95% confidence intervals. At the end of each section, I will use a bar graph to display the results from the three-way interaction, which provide an additional perspective on the role of each of the independent variables.

5.2.3.1. Group, ExpectedMood and MoodSelectType

In this model, the binary dependent variable is subjunctive production, coded as "1" for subjunctive production and "0" for indicative production, and the independent variables are Group (SDC, LCI, AdvHS or IntHS), MoodSelectType (Intensional or Polarity) and ExpectedMood (Subjunctive or Indicative). (Note that all variable names are capitalized

throughout the remainder of Section 5.2.) The goal of GLMM #1 is to determine whether Group, MoodSelectType and ExpectedMood condition participants' likelihood of producing subjunctive mood. In order to focus on these three variables, GLMM #1 does not take into account PrimeStatus or Frequency. The role of these critical variables will be addressed in GLMM #2 and #3, respectively.

Before running the model, however, I had to manually introduce variability into each of the invariant experimental conditions, since GLMMs cannot run with totally invariant data. The SDCs were invariant in the Intensional Subjunctive (100% subjunctive production) and Intensional Indicative (0% subjunctive production) conditions while the LCIs were invariant in the Intensional Subjunctive (100% subjunctive production) condtion only. To fix this lack of variance, I selected two items in each of these conditions (one from Version A, one from Version B) and changed participants' responses from either Indicative to Subjunctive (Intensional Indicative condition) or Subjunctive to Indicative (Intensional Subjunctive condition). This alteration, which affected only 0.2% of the CEPT data, makes it possible to run the necessary statistical models.

5.2.3.1.1. Main effects

A GLMM revealed significant main effects of Group ($F(3,124) = 14.293$, $p < .001$), MoodSelectType ($F(1,455) = 12.902$, $p < .001$, OR = 3.11) and ExpectedMood ($F(1,118) = 244.783$, $p < .001$, OR = 471.54). With respect to Group, pairwise comparisons revealed that the SDCs' (vs. AdvHS: $p < .01$, OR = 7.78; vs. IntHS: $p < .001$, OR = 13.45) and LCIs' (vs. AdvHS: $p < .001$, OR = 19.63; vs. IntHS: $p < .001$, OR = 33.95) overall odds of producing subjunctive were many times greater than the odds for the two HS groups.

Table 20.
Odds of using subjunctive by Group

Group	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
SDC	0.793	0.470	2.210	68.8%	46.6%	84.8%
LCI	1.719	0.525	5.579	84.8%	66.5%	94.0%
AdvHS	-1.258	0.433	0.284	22.1%	10.8%	40.1%
IntHS	-1.806	0.385	0.164	14.1%	7.1%	26.2%

Table 20 presents each group's overall predicted probability of subjunctive production in the CEPT. Because this is a main effect, however, it does not take into consideration the critical factors MoodSelectType and ExpectedMood. Consequently, the takeaway from Table 20 is that the SDCs and LCIs, not surprisingly, are more likely (in general) to produce subjunctive forms.

Pairwise comparisons for MoodSelectType (Table 21) revealed that overall, participants' odds of producing subjunctive were over three times higher with Polarity than with Intensional mood selection, $p < .001$, OR = 3.11. On the surface, this finding is surprising, given that HSs are hypothesized to produce more subjunctive with Intensional mood selection. However, as we will see later in Section 5.2.3.1.2, this effect is driven by the tendency for participants in all groups to "overproduce" subjunctive more with Polarity than with Intensional mood selection.

Table 21.
Odds of using subjunctive by MoodSelectType

M. Select	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
Polarity	0.430	0.258	1.537	60.6%	48.0%	72.0%
Intensional	-0.706	0.330	0.494	33.0%	20.5%	48.6%

Finally, pairwise comparisons for ExpectedMood (Table 22) revealed that participants' odds of producing subjunctive when Subjunctive was expected were significantly higher than when Indicative was expected, $p < .001$, OR = 471.54. It is worth

noting here that the extremely high effect size (OR = 471.54) is largely driven by the performance of the Spanish-dominant groups.

Table 22.
Odds of using subjunctive by ExpectedMood

Exp. Mood	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
Subjunctive	2.940	0.304	18.916	95.0%	91.2%	97.2%
Indicative	-3.216	0.333	0.040	3.9%	2.0%	7.2%

Summary of main effects

In summary, the GLMM revealed three statistically significant main effects. First, Spanish-dominant groups (especially the LCIs) have higher overall odds of producing subjunctive than the HSs. Second, participants (as a whole) produce significantly more subjunctive with Polarity, as opposed to Intensional, mood selection. Finally, participants (as a whole) produce significantly more subjunctive when subjunctive mood is expected.

In order to better understand these effects, we now explore the two-way interactions between Group, MoodSelectType and ExpectedMood, respectively.

5.2.3.1.2. Two-way interactions

The GLMM also revealed statistically significant two-way interactions between Group and ExpectedMood ($F(3,2740) = 35.583$, $p < .001$), MoodSelectType and ExpectedMood ($F(1,455) = 36.239$, $p < .001$), and Group and MoodSelectType ($F(3,2740) = 2.856$, $p < .05$). In the remainder of Section 5.3.3.2, we will discuss the first two interactions only, since only these interactions shed light on the RQs of the present study[26].

[26] The two-way interaction between Group and MoodSelectType reveals whether the participant groups differ from one another in their tendency to produce more subjunctive with Intensional or Polarity mood selection, respectively. Because this interaction does not take into account Expected Mood, however, its interpretation is not relevant here.

In Section 5.2.3.1.1, we noted a significant main effect of Group on participants' overall likelihood of producing subjunctive. The Group x ExpectedMood interaction, demonstrated in Table 23, clarifies this main effect by revealing whether these group differences surface in both ExpectedSubjunctive and ExpectedIndicative conditions.

Table 23.
Odds of using subjunctive: Group x ExpectedMood

Group	E Mood	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
SDC	Subjunctive	4.995	0.602	147.673	99.3%	97.8%	99.8%
	Indicative	-3.410	0.577	0.033	3.2%	1.0%	9.3%
LCI	Subjunctive	5.907	0.780	367.602	99.7%	98.8%	99.9%
	Indicative	-2.469	0.555	0.085	7.8%	2.8%	20.2%
AdvHS	Subjunctive	1.465	0.371	4.328	81.2%	67.4%	90.1%
	Indicative	-3.980	0.654	0.019	1.8%	0.5%	6.3%
IntHS	Subjunctive	-0.608	0.378	0.544	35.3%	20.4%	53.6%
	Indicative	-3.003	0.477	0.050	4.7%	1.9%	11.3%

An additional GLMM revealed a significant effect of Group in the ExpectedSubjunctive conditions ($F(3,148) = 34.624$, $p < .001$). When subjunctive is expected, the SDCs (vs. AdvHS: $p < .001$, OR = 34.12; vs. IntHS: $p < .001$, OR = 271.24) and the LCIs (vs. AdvHS: $p < .001$, OR = 84.94; vs. IntHS: $p < .001$, OR = 675.19) have significantly higher odds of producing subjunctive as compared to the HS groups. Within the HS groups, odds of producing subjunctive were modulated by proficiency: The AdvHSs' odds of producing subjunctive in the ExpectedSubjunctive conditions were higher than the odds for the IntHS, $p < .001$, OR = 7.95.

In the ExpectedIndicative condition, on the other hand, there was no significant effect of Group, $F(3,243) = 1.279$, $p > .2$, suggesting that HSs diverge from the Spanish-dominant groups only when subjunctive mood forms are expected.

To complete our analysis of this interaction effect, however, we must shift our focus from between-group comparisons, which highlight the ways in which HSs differ from Spanish-dominant speakers, to *within-group* comparisons, which underscore the systematic nature of the heritage grammars themselves. An additional GLMM showed that the SDCs ($F(1,1104) = 139.352$, $p < .001$, OR = 4865.87), LCIs ($F(1,1881) = 95.995$, $p < .001$, OR = 4341.61), AdvHS ($F(1,714) = 77.941$, $p < .001$, OR = 231.60), and IntHS ($F(1,109) = 39.057$, $p < .001$, OR = 10.97) all showed significantly higher odds of producing subjunctive in the ExpectedSubjunctive conditions. This finding is not surprising, especially for the first three groups. It is noteworthy, however, the extent to which the IntHSs, whose raw proportion of subjunctive production was less than 40% overall, display strong, systematic sensitivity to ExpectedMood. Pairwise comparisons reveal that the IntHSs' odds of producing subjunctive mood are over ten times higher in the ExpectedSubjunctive conditions than in ExpectedIndicative conditions. We would not expect to find this if IntHSs were performing randomly with subjunctive mood.

The interaction between MoodSelectType and ExpectedMood (Table 24) was also statistically significant, meaning that the impact of MoodSelectType on participants' subjunctive production was modulated by ExpectedMood. Additional GLMMs revealed that in ExpectedSubjunctive conditions, participants' odds of producing subjunctive were slightly higher with Intensional than with Polarity selection, $F(1,157) = 2.531$, $p > .1$, OR = 2.16. This marginally significant tendency was driven by the HS groups, who were more likely to produce subjunctive with Intensional mood selection (See Section 5.2.3.1.3).

In ExpectedIndicative conditions, however, participants' odds of producing subjunctive were significantly higher with Polarity mood selection, $F(1,2740) = 55.436$, p

< .001, OR = 20.93. This interaction confirms the observation that participants' odds of "overproducing" subjunctive with ExpectedIndicative items are significantly higher with Polarity mood selection, meaning that they are over twenty times more likely to produce subjunctive in [+presuppositional] contexts than after the complementizer *porque*.

Table 24.
Odds of using subjunctive: ExpectedMood x MoodSelectType

E. Mood	Select Type	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
Subjunctive	Intensional	3.324	0.416	27.771	96.5%	92.4%	98.4%
	Polarity	2.555	0.359	12.871	92.8%	86.4%	96.3%
Indicative	Intensional	-4.736	0.467	0.009	0.9%	0.3%	2.2%
	Polarity	-1.695	0.294	0.184	15.5%	9.3%	24.8%

Summary of two-way interactions

In summary, participants' odds of producing subjunctive are significantly affected by Group, but only in the ExpectedSubjunctive conditions. When considered individually, however, all groups, including the IntHSs, are significantly more likely to produce subjunctive in ExpectedSubjunctive as opposed to ExpectedIndicative conditions. Finally, participants' odds of producing subjunctive are conditioned by MoodSelectType, meaning that they are more likely to "overproduce" subjunctive with Polarity, as opposed to Intensional, selection. To complete our analysis of the combined effects of Group, MoodSelectType and ExpectedMood on participants' subjunctive production, we now turn to the three-way interaction effect, which presents the most detailed picture of the data.

5.2.3.1.3. Three-way interaction

GLMM #1 did not reveal a significant effect of the Group x MoodSelectType x ExpectedMood interaction, $F(3,2740) = 1.982$, $p > .1$. Nonetheless, further analyzing the

three-way interaction, shown below in Table 25, allows us to make a number of critically important statistical comparisons which will guide our analysis moving forward.

To begin, we examine the impact of Group on subjunctive production in each of the four experimental conditions: IntensionalSubjunctive, IntensionalIndicative, PolaritySubjunctive and PolarityIndicative. In other words, do the groups differ from one another in each condition?

Table 25.
Odds of using subjunctive: Group x MoodSelectType x ExpectedMood

Group	M. Select	E. Mood	Logodds	SE	Odds	Probability	CI-Lower	CI-Upper
SDC	Intensional	subj	5.236	0.825	187.917	99.5%	97.4%	99.9%
		indic	-4.703	0.853	0.009	0.9%	0.2%	4.6%
	Polarity	subj	4.755	0.714	116.164	99.1%	96.6%	99.8%
		indic	-2.117	0.505	0.120	10.7%	4.2%	24.6%
LCI	Intensional	subj	6.298	1.127	543.484	99.8%	98.4%	100.0%
		indic	-4.623	0.841	0.010	1.0%	0.2%	4.9%
	Polarity	subj	5.516	0.871	248.638	99.6%	97.8%	100.0%
		indic	-0.315	0.463	0.730	42.2%	22.6%	64.6%
AdvHS	Intensional	subj	1.866	0.422	6.462	86.6%	73.7%	93.7%
		indic	-5.958	1.100	0.003	0.3%	0.0%	2.2%
	Polarity	subj	1.063	0.410	2.895	74.3%	56.2%	86.7%
		indic	-2.002	0.455	0.135	11.9%	5.2%	24.9%
IntHS	Intensional	subj	-0.103	0.417	0.902	47.4%	28.2%	67.4%
		indic	-3.659	0.594	0.026	2.5%	0.8%	7.6%
	Polarity	subj	-1.113	0.420	0.329	24.7%	12.5%	43.1%
		indic	-2.347	0.494	0.096	8.7%	3.5%	20.2%

In the IntensionalSubjunctive condition, there was a statistically significant difference in the groups' odds of producing subjunctive, $F(3,236) = 19.957$, $p < .001$. Pairwise comparisons revealed that the SDCs (vs. AdvHS: $p < .001$, OR = 29.08; vs. IntHS: $p < .001$, OR = 214.65) and LCIs (vs. AdvHS: $p < .001$, OR = 84.10; vs. IntHS: $p < .001$, OR = 602.45) had significantly higher odds of producing subjunctive than the heritage groups. While there was no statistically significant difference between the Spanish-

dominant groups ($p > .4$, OR = 2.89), the AdvHS' odds of producing subjunctive were over 7 times higher than the odds for the IntHS ($p < .001$, OR = 7.16), meaning that proficiency modulates HSs' likelihood of producing Intensional Subjunctive.

In the IntensionalIndicative condition, where indicative is expected, there was no statistically significant effect of Group, $F(3,1098) = 1.346$, $p > .2$, meaning that the groups did not differ from one another in their odds of producing subjunctive in this condition, all pairwise p's > .05. However, the IntHSs showed a marginally significant tendency ($p = .059$, OR = 9.96) to produce more subjunctive than the AdvHSs in this condition. In a way, it is surprising that the IntHSs are more likely to produce subjunctive in this condition[27]. Nonetheless, it is likely that their higher production of subjunctive reflects instability in lexical knowledge (e.g., classifying –AR verbs as –ER/-IR) rather than a legitimate tendency to overextend subjunctive mood morphology.

Not surprisingly, huge group differences emerge in the PolaritySubjunctive condition, $F(3,197) = 29.954$, $p < .001$. Pairwise comparisons revealed that the SDCs (vs. AdvHS: $p < .001$, OR = 40.13; vs. IntHS: $p < .001$, OR = 353.54) and the LCIs (vs. AdvHS: $p < .001$, OR = 85.88; vs. IntHS: $p < .001$, OR = 756.73) had vastly higher odds of producing subjunctive in this condition when compared to the HS groups. As in the IntensionalSubjunctive condition, there were no differences between the Spanish-dominant groups ($p > .4$, OR = 2.14). There were, however, proficiency-modulated differences between the HS groups, as the AdvHS' ($p < .001$, OR = 8.81) odds of producing subjunctive were over eight times higher than the odds for the IntHS.

---

[27] Participant #54, an IntHS, produced two of the six total instances of subjunctive after *porque*. On Item #35, she said "Busco una tableta porque mi mamá siempre [le-a_{SUBJ}] en los aeropuertos" and on Item #49, she said "Busco un par de zapatillas porque muchas veces mi papá s[sa-ke_{SUBJ}] la basura y acaba con mugre en los pies."

In the PolarityIndicative condition, we begin to see emerging differences between the LCIs and the other experimental groups. Unlike in the IntensionalIndicative condition, where there was no effect of Group, a GLMM in this condition revealed significant group differences, $F(3,130) = 4.726$, $p < .01$. Pairwise comparisons revealed a surprising tendency for the LCIs to produce subjunctive mood here, in spite of the clear presupposition in the experimental context. The LCIs' (vs. SDC, $p < .01$, OR = 6.06; vs. AdvHS: $p < .01$, OR = 5.40; vs. IntHS: $p < .01$, OR = 7.63) odds of producing subjunctive in this condition were much higher than the odds for the other groups (i.e., at least 5.4 times higher odds for all comparisons), a pattern we will also see in Section 5.3.

While between-group comparisons are important for the present analysis, even more important are comparisons across conditions *within* each participant group. The first within-group comparison that we consider here is the role of MoodSelectType. Do each of the groups' odds of producing subjunctive change depending on the type of mood selection in a given item? If so, do these differences in odds appear in both IndicativeExpected and SubjunctiveExpected conditions?

When Subjunctive is expected, the SDCs ($F(1,2427) = .249$, $p > .6$, OR = 1.62) and the LCIs ($F(1,2740) = .378$, $p > .5$, OR = 2.19) showed a slight tendency to produce more subjunctive with Intensional selection. The AdvHS ($F(1,58) = 4.561$, $p < .05$, OR = 2.23) and the IntHS ($F(1,49) = 7.869$, $p < .01$, OR = 2.75), however, exhibited a much stronger MoodSelectType effect, demonstrating significantly higher odds of producing subjunctive with Intensional, as opposed to Polarity mood selection.

When Indicative is expected, each of the groups show the same pattern, albeit to differing extents. The SDCs ($F(1,2740) = 10.526$, $p < .01$, OR = 13.28), LCIs (F(1,2740)

= 30.424, $p < .001$, OR = 74.29), AdvHS (F(1,2740) = 13.931, $p < .001$, OR = 52.25) and IntHS (F(1,2740) = 6.074, $p < .05$, OR = 3.71) all had significantly higher odds of producing Subjunctive with Polarity, as opposed to Intensional mood selection. These intra-group comparisons confirm that all groups overgeneralize subjunctive more with Polarity than Intensional mood selection.

Thus far, the pairwise comparisons examined here reflect that the SDCs and LCIs behave (mostly) as expected, the AdvHS perform similarly (if also less accurately) and the IntHS perform in a qualitatively distinct manner. The final within-group comparison, which examines whether each individual group is sensitive to ExpectedMood in each of the two MoodSelect conditions, suggests that all groups, including the IntHS, show sensitivity to ExpectedMood when producing subjunctive mood morphology in Spanish.

With Intensional mood selection, the SDCs (F(1,2740) = 80.158, $p < .001$, OR = 20,879.67), LCIs (F(1,2740) = 65.907, $p < .001$, OR = 54,348.40), AdvHS (F(1,2740) = 50.496, $p < .001$, OR = 2154.00) and IntHS (F(1,267) = 41.284, $p < .001$, OR = 34.69) all are substantially more likely to produce subjunctive with ExpectedSubjunctive items. Similarly, with Polarity mood selection, the SDCs (F(1,1854) = 83.607, $p < .001$, OR = 968.03), LCIs (F(1,1430) = 44.361, $p < .001$, OR = 340.60), AdvHS (F(1,94) = 51.871, $p < .001$, OR = 21.44) and IntHS (F(1,109) = 7.846, $p < .01$, OR = 3.43) are all more likely to produce subjunctive with ExpectedSubjunctive items. It is noteworthy that all groups, including the HSs, are sensitive to ExpectedMood with both types of mood selection. The three-way interaction effect is represented graphically in Figure 9.

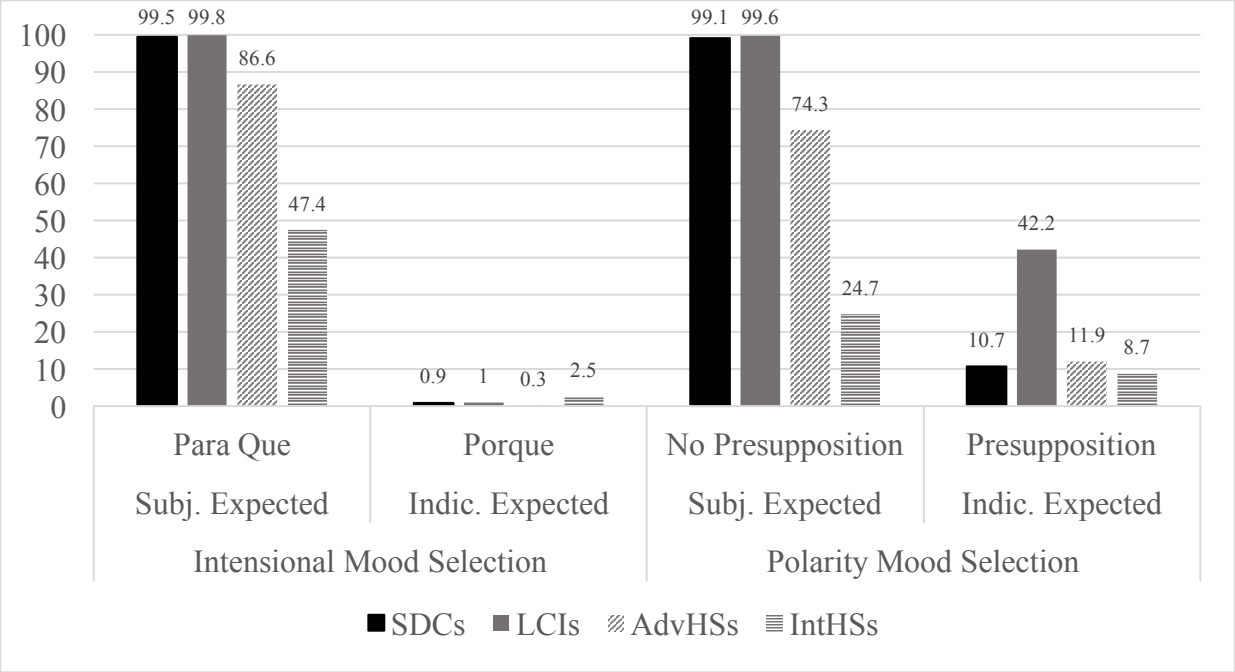

Figure 9: Probability of subjunctive use: Group x MoodSelectType x ExpectedMood

Summary of three-way interactions

An analysis of the three-way interaction reveals a few important patterns. First, there were significant differences between the experimental groups across experimental conditions. In both ExpectedSubjunctive conditions, the HS groups were significantly less likely than the SDCs and LCIs to produce subjunctive mood forms. In the ExpectedIndicative conditions, however, the HS groups did not diverge from the SDCs, suggesting that HS divergence from controls tends to be localized in ExpectedSubjunctive contexts. Only the LCIs, who showed a curious tendency to produce subjunctive in PolarityIndicative contexts, diverged from the other experimental groups with ExpectedIndicative items. Second, there were important within-group comparisons which highlight the critical role of MoodSelection for each group. In the ExpectedSubjunctive items, the HSs were more likely to produce subjunctive with Intensional as opposed to Polarity mood selection, showing an Intensional facilitation effect not present for the SDCs

or LCIs. In ExpectedIndicative items, however, all groups patterned similarily, producing significantly more subjunctive with Polarity, as opposed to Intensional mood selection. Finally, despite immense between-group differences in odds of subjunctive production, all groups displayed sensitivity to ExpectedMood, producing significantly more subjunctive when Subjunctive was expected with both Intensional and Polarity Mood selection.

5.2.3.2. Group, MoodSelectType and Frequency

The goal of GLMM #2 is to examine the extent to which the relative frequency of different verbs modulates participants' likelihood of producing subjunctive mood. In order to model this potential frequency effect, GLMM #2 focuses exclusively on experimental items where subjunctive mood is expected, reducing the total number of data points from 2756 to 1855. The independent variables in GLMM #2, therefore, are Group (SDC, LCI, AdvHS or IntHS), MoodSelectType (Intensional or Polarity) and Frequency (Frequent or Infrequent). Frequency, as explained in Section 4.4.2, was determined by dividing the twelve verbs in each subjunctive type (intensional and polarity) into two groups: Frequent (the six most frequent, according to Davies' (2006) frequency dictionary) and Infrequent (the six least frequent, according to Davies.)

As in GLMM #1, slight alterations were made to the data in order to introduce variation into otherwise invariant conditions. The LCIs produced subjunctive 100% of the time with Frequent verbs and Intensional mood selection. Consequently, I manually selected one of the items in this condition and changed the participant's response from "1" (produced subjunctive) to "0" (produced indicative). This change, which affected .05% of the data, allowed GLMM #2 to run.

5.2.3.2.1. Main effects

GLMM #1 analyzed the effects of Group and MoodSelectType. To avoid redundancy, I will refrain from discussing the main effects of these variables in GLMM #2, limiting my discussion to the role of Frequency. GLMM #2 revealed a non-significant main effect of Frequency, ($F(1,85) = 1.104$, $p > .2$, OR = 1.57). Pairwise comparisons showed that overall, participants' odds of producing subjunctive mood forms were slightly, though not significantly, higher (OR = 1.57) with Frequent as opposed to Infrequent verbs.

Summary of main effects

Overall, participants are not more likely to produce subjunctive mood with Frequent verb forms. In order to determine whether this is true for all experimental groups and with both types of mood selection, we now turn to the two-way interaction effects.

5.2.3.2.2 Two-way interactions

GLMM #2 revealed a non-significant interaction between Group and Frequency ($F(3,1839) = 1.519$, $p > .2$). Still, it is important for us to further examine this interaction to determine (a) whether participants in each group are more likely to produce subjunctive with Frequent verbs and (b) whether the HS groups diverge more from the SDCs and LCIs with Frequent or Infrequent verbs, respectively.

Pairwise comparisons revealed that neither the SDCs ($F(1,1839) = .091$, $p > .7$, OR = 0.74) nor the LCIs ($F(1,1839) = .010$, $p > .9$, OR = 1.11) were significantly more likely to produce subjunctive with Frequent verbs. The heritage groups, however, were quite sensitive to Frequency. Both the AdvHS ($F(1,48) = 11.833$, $p < .01$, OR = 3.64) and the IntHS ($F(1,35) = 4.342$, $p < .05$, OR = 2.06) had significantly higher odds of producing

subjunctive with Frequent verbs, revealing that their variable production of subjunctive mood is modulated, at least to some extent, by lexical frequency.

When we focus on the between-group perspective of this interaction, we see that both HS groups diverge less from the SDCs with Frequent verbs. Not surprisingly, the SDCs' odds of producing subjunctive with Frequent verbs were significantly higher than the odds for the AdvHS ($p < .01$, OR = 13.41) and the IntHS ($p < .001$, OR = 172.26). With Infrequent verbs, the SDCs' odds of producing subjunctive were still significantly higher than the odds for the AdvHSs ($p < .001$, OR = 65.50) and the IntHSs ($p < .001$, OR = 476.28). Notably though, the HS groups' odds of producing subjunctive diverged more sharply from the SDCs' odds with Infrequent verbs than with Frequent verbs, suggesting again that verb Frequency modulates the extent to which HSs diverge diverge from Spanish-dominant control groups in subjunctive production.

GLMM #2 also revealed a non-significant interaction between Frequency and MoodSelectType, $F(1,85) = 0.415$, $p > .5$, meaning that Frequency did not impact subjunctive production significantly more with Intensional, as opposed to Polarity mood selection. For both Intensional ($F(1,93) = 1.370$, $p > .2$, OR = 2.08) and Polarity ($F(1,77) = 0.087$, $p > .7$, OR = 1.19) mood selection, participants' odds of subjunctive production were not significantly higher with Frequent verbs than with Infrequent verbs.

Summary of two-way interactions

Neither of the two-way interaction effects involving Frequency was statistically significant. Nonetheless, pairwise comparisons revealed small impacts of Frequency in both within-group and between-group comparisons. Both HS groups, unlike the SDCs and LCIs, were significantly more likely to produce subjunctive with Frequent verbs. Not

surprisingly then, the HS groups' odds of subjunctive production differed more from the Spanish-dominant groups with Infrequent verbs than with Frequent verbs. Finally, the effect of Frequency on participants' odds of subjunctive production was not different with different types of mood selection.

5.2.3.2.3. Three-way interactions

The three-way interaction of Group x MoodSelectType x Frequency (Table 8) was also not statistically significant, $F(3,1839) = 0.050$, $p > .9$. We will examine the interaction in greater detail, however, because it allows us to make critical between- and within-group comparisons across all possible combinations of experimental conditions and groups.

To begin our exploration of this interaction, we will examine between-group differences in each of the four experimental conditions: IntensionalFrequent, IntensionalInfrequent, PolarityFrequent and PolarityInfrequent.

Table 26.
Odds of using subjunctive: Group x MoodSelectType x Frequency

Group	M. Select	Freq	Logodds	SE	Odds	Probability	CI-Lower	CI-Upper
SDC	Intensional	Freq	5.130	1.138	169.017	99.4%	94.8%	100.0%
		Infreq	5.087	1.135	161.903	99.4%	94.6%	100.0%
	Polarity	Freq	4.430	0.886	83.931	98.8%	93.6%	99.8%
		Infreq	5.062	1.135	157.906	99.4%	94.5%	100.0%
LCI	Intensional	Freq	5.208	1.167	182.728	99.5%	94.9%	100.0%
		Infreq	5.099	1.163	163.858	99.4%	94.4%	100.0%
	Polarity	Freq	5.250	1.166	190.566	99.5%	95.1%	100.0%
		Infreq	5.146	1.163	171.743	99.4%	94.6%	100.0%
AdvHS	Intensional	Freq	2.846	0.570	17.219	94.5%	84.7%	98.2%
		Infreq	1.184	0.516	3.267	76.6%	53.9%	90.1%
	Polarity	Freq	1.521	0.525	4.577	82.1%	61.7%	92.9%
		Infreq	0.603	0.506	1.828	64.6%	40.0%	83.4%
IntHS	Intensional	Freq	0.398	0.524	1.489	59.8%	34.3%	80.9%
		Infreq	-0.728	0.525	0.483	32.6%	14.5%	57.9%
	Polarity	Freq	-1.136	0.525	0.321	24.3%	10.1%	47.8%
		Infreq	-1.453	0.529	0.234	19.0%	7.5%	40.2%

With Intensional selection and Frequent verbs, there was a statistically significant effect of Group, $F(3,282) = 10.052$, $p < .001$. Pairwise comparisons showed that the SDCs ($p < .001$), LCIs ($p < .001$) and AdvHS ($p < .01$) all had significantly higher odds of producing subjunctive than the IntHS. Interestingly though, the SDCs' ($p = .063$, OR = 9.82) and the LCIs' ($p = .060$, OR = 10.61) odds of producing subjunctive in this condition were only marginally higher than the odds for the AdvHS, whose predicted probability of subjunctive production (94.5%) nearly reached ceiling.

With Intensional selection and Infrequent verbs, however, larger group differences emerged, $F(3,235) = 13.406$, $p < .001$. This time, the SDCs' (vs. AdvHS: $p < .01$, OR = 49.55; vs. IntHS: $p < .001$) and the LCIs' (vs. AdvHS: $p < .01$, OR = 50.15; vs. IntHS: $p < .001$) odds of producing subjunctive were significantly higher than the odds for either of the HS groups. By comparing the OR's in this condition with the OR's from the previous condition, we gain a valuable insight into the role of frequency in group differences between the HSs and Spanish-dominant groups. To illustrate this point, consider the comparison between the AdvHSs and the SDCs and LCIs. With Frequent verbs, the SDCs' and LCIs' odds of producing subjunctive were approximately ten times higher than the odds for the AdvHS. With Infrequent verbs, however, the odds for the Spanish-dominant groups were over *fifty* times higher than the odds for the AdvHS, whose predicted probability of producing subjunctive was just 76.6% in this condition. Clearly, Frequency affects the extent to which the HSs diverge from their Spanish-dominant counterparts.

A similar pattern emerges in the Polarity conditions, where the Spanish-dominant groups once again perform in a far more target-like manner than the HSs. With Polarity

selection and Frequent verbs, there was a statistically significant effect of Group, $F(3,225) = 17.263$, $p < .001$. Pairwise comparisons revealed that the SDCs (vs. AdvHS: $p < .01$, OR = 18.34; vs. IntHS: $p < .001$) and the LCIs (vs. AdvHS: $p < .01$, OR = 41.64; vs. IntHS: $p < .001$) had significantly higher odds of producing subjunctive mood than either of the heritage groups, who once again diverged from each other by proficiency. While the AdvHS' odds of producing subjunctive are 18 times less than the odds of the SDCs and 40 times less than the odds of the LCIs, their predicted probability of 82.1% suggests high accuracy in this condition with Frequent verbs.

With Polarity selection and Infrequent verbs, the same group differences emerge, $F(3,228) = 16.844$, $p < .001$: the SDCs' (vs. AdvHS: $p < .001$, OR = 86.40; vs. IntHS: $p < .001$) and LCIs' (vs. AdvHS: $p < .001$, OR = 93.97; vs. IntHS: $p < .001$) odds of producing subjunctive are significantly higher than the odds for the heritage groups. What differs with Infrequent verbs is the *magnitude* of the group differences, as expressed by the ORs. With Frequent verbs, the SDCs' odds of producing subjunctive were 18 times higher than the odds for the AdvHS. With Infrequent verbs, however, the SDCs' odds are over *ninety* times higher than the odds for the AdvHS, whose predicted probability of producing subjunctive drops to 64.6% in this condition. This staggering, five-fold increase in the difference between the SDCs and AdvHS provides additional evidence that Frequency affects the extent to which HSs diverge from Spanish-dominant control groups.

Recall from Section 5.2.3.2.2 that both the AdvHSs and IntHSs, but not the SDCs and LCIs, were significantly more likely to produce subjunctive mood with Frequent verbs. By examining another perspective of the three-way interaction, we can determine whether this apparent frequency effect is conditioned by mood selection type.

With Intensional mood selection, there was no significant effect of Frequency for either the SDCs (F(1,1839) = 0.001, $p > .9$, OR = 1.04) or the LCIs (F(1,1839) = 0.005, $p > .9$, OR = 1.11). For the AdvHS (F(1,56) = 9.081, $p < .01$, OR = 5.27) and the IntHS (F(1,36) = 5.245, $p < .05$, OR = 3.08), however, the odds of producing subjunctive with Intensional mood selection were between 3 and 5 times higher with Frequent verbs. With Polarity mood selection, there was again no significant effect of Frequency for the SDCs (F(1,1618) = 0.241, $p > .6$, OR = 0.53) and LCIs (F(1,1839) = 0.005, $p > .9$, OR = 1.11). Curiously, however, there was also no significant effect of Frequency for the AdvHS (F(1,37) = 3.415, $p = .073$, OR = 2.50) and the IntHS (F(1,34) = 0.426, $p > .5$, OR = 1.37), though both groups showed very slight tendencies to produce more subjunctive with Frequent verbs. It is unclear why Frequency effects in the HSs' production of subjunctive are limited to items with Intensional, rather than Polarity, mood selection.

The three-way interaction effect is represented graphically in Figure 10.

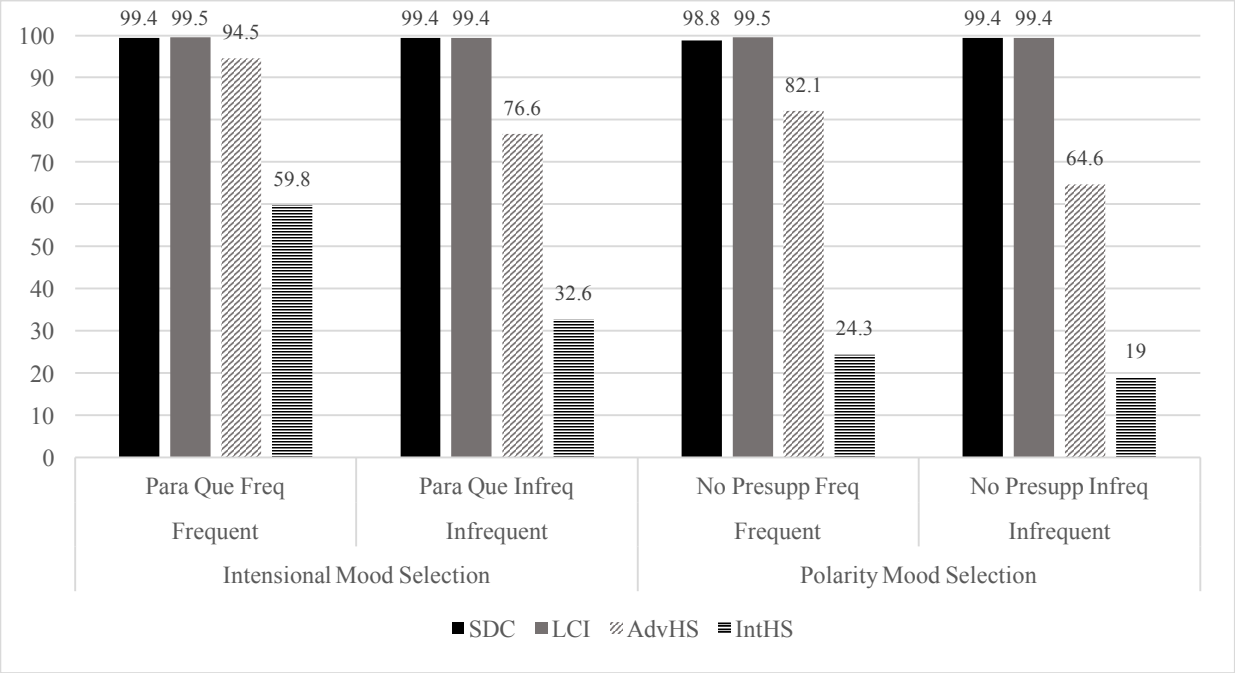

Figure 10: Probability of subjunctive use: Group x MoodSelectType x Frequency

Summary of three-way interactions

Analysis of the three-way interaction effect reveals illustrative between- and within-group differences. While the HS groups differed from the Spanish-dominant groups in each of the four experimental conditions, they diverged much more from controls with Infrequent verbs than with Frequent verbs, a trend that held for both Intensional and Polarity mood selection. Despite this finding, within-group comparisons revealed that the effect of Frequency was only significant for the HSs with Intensional mood selection.

5.2.3.3. Group, MoodSelectType and PrimeStatus

The goal of GLMM #3 is to examine the extent to which the structural priming modulates participants' likelihood of producing subjunctive mood. In order to model this potential priming effect, GLMM #3 focuses exclusively on experimental items where subjunctive mood is expected. The independent variables in GLMM #3, therefore, are Group (SDC, LCI, AdvHS or IntHS), MoodSelectType (Intensional or Polarity) and PrimeStatus (Prime or Baseline).

As in GLMMs #1 and #2, slight alterations were made to the data to introduce variation into otherwise invariant conditions. The SDCs produced subjunctive 100% of the time in the (a) IntensionalBaseline, (b) IntensionalPrime and (c) PolarityPrime conditions, respectively, while the LCIs produced subjunctive 100% of the time in the (a) IntensionalBaseline and (b) PolarityPrime conditions. For each of the invariant cells, I changed two responses from "1" (subjunctive produced) to "0" (indicative produced). In total, these changes affected a total of 0.5% of the total data that went into this model.

5.2.3.3.1. Main effects

GLMM #1 analyzed the effects of Group and MoodSelectType. To avoid redundancy, I will refrain from discussing the main effects of these variables in GLMM #3, limiting my discussion to the role of PrimeStatus.

GLMM #3 revealed a non-significant main effect of PrimeStatus, ($F(1,1839) = 0.330$, $p > .5$. Pairwise comparisons revealed that participants' odds of producing subjunctive mood were not significantly higher in the Prime condition as opposed to the Baseline condition, $p > .5$, OR = 1.17.

Summary of main effects

Overall, participants are not more likely to produce subjunctive mood morphology in the Prime, as opposed to the Baseline conditions.

5.2.3.3.2. Two-way interactions

In Section 5.2.3.3.1, we saw that there was no significant effect of PrimeStatus on participants' likelihood of producing subjunctive mood. It is possible, however, that an effect of priming emerges either within a specific group or within a certain type of mood selection. To explore these possibilities, this section analyzes the two-way interactions between (a) PrimeStatus and Group and (b) PrimeStatus and MoodSelectType.

GLMM #3 revealed a non-significant effect of Group x PrimeStatus ($F(3,1839) = 0.315$, $p > .8$. Nonetheless, exploring pairwise group comparisons offers us the opportunity to test whether the effect of priming impacted the experimental groups differently. Additional GLMMs revealed that for the SDCs ($F(1,1839) = 0.099$, $p > .7$, OR = 1.24), the LCIs ($F(1,1839) = 0.279$, $p > .5$, OR = 1.53), the AdvHS($F(1,1839) = 0.290$, $p > .5$, OR =

0.87) and the IntHS ($F(1,1839) = 0.333$, $p > .5$, OR = 1.15), the odds of producing subjunctive mood were not significantly higher in the Prime condition than in the Baseline condition. It's possible, however, that the lack of an interaction effect here is caused by effects of MoodSelectType on the effect of priming. In order to rule out this possibility, we next examine the interaction between MoodSelectType and PrimeStatus.

GLMM #3 also revealed a non-significant effect of PrimeStatus x MoodSelectType ($F(1,1839) = 0.001$, $p > .9$. Additional GLMMs showed that with there was no effect of PrimeStatus in the Intensional ($F(1,1839) = 0.130$, $p > .7$, OR = 1.17) or Polarity ($F(1,1839) = 0.213$, $p > .6$, OR = 1.18) mood selection condition.

Summary of two-way interactions

Analyses of the two-way interactions involving PrimeStatus reveal that PrimeStatus does not impact odds of subjunctive production for any of the experimental groups, with either Intensional or Polarity mood selection, respectively.

5.2.3.3.3 Three-way interaction

Up to this point, there is no reason to believe that PrimeStatus impacts odds of producing subjunctive (a) for any of the experimental groups and (b) in either mood selection type. To rule out the possibility that PrimeStatus plays a more fine-grained role in participants' odds of subjunctive production, we now examine the three-way interaction which has the power to reveal whether PrimeStatus plays a significant role for *any* of the groups in *any* experimental condition.

The GLMM revealed that the Group x MoodSelect x PrimeStatus interaction (Table 27) was not statistically significant, ($F(3,1839) = 0.170$, $p > .9$).

Table 27.
Odds of using subjunctive: Group x MoodSelectType x PrimeStatus

Group	M. Select	Prime Status	Logodds	SE	Odds	Probability	CI-Lower	CI-Upper
SDC	Intensional	Prime	4.467	0.849	87.095	98.9%	94.3%	99.8%
		Base	4.474	0.859	87.707	98.9%	94.2%	99.8%
	Polarity	Prime	4.379	0.847	79.758	98.8%	93.8%	99.8%
		Base	3.936	0.742	51.213	98.1%	92.3%	99.5%
LCI	Intensional	Prime	5.229	1.127	186.606	99.5%	95.3%	100.0%
		Base	4.354	0.858	77.789	98.7%	93.5%	99.8%
	Polarity	Prime	4.360	0.858	78.257	98.7%	93.6%	99.8%
		Base	4.381	0.861	79.918	98.8%	93.6%	99.8%
AdvHS	Intensional	Prime	1.768	0.487	5.859	85.4%	69.0%	93.9%
		Base	2.036	0.497	7.660	88.5%	74.1%	95.3%
	Polarity	Prime	1.023	0.471	2.782	73.6%	52.2%	87.6%
		Base	1.038	0.470	2.824	73.8%	52.6%	87.8%
IntHS	Intensional	Prime	-0.147	0.483	0.863	46.3%	24.8%	69.3%
		Base	-0.163	0.481	0.850	45.9%	24.6%	68.8%
	Polarity	Prime	-1.132	0.484	0.322	24.4%	11.0%	45.8%
		Base	-1.400	0.484	0.247	19.8%	8.6%	39.2%

To determine whether PrimeStatus played a significant (within-group) role with (a) Intensional or (b) Polarity mood selection, additional GLMMs were run.

With Intensional mood selection, there was no significant effect of PrimeStatus for the SDCs ($F(1,1839) = 0.000$, $p > .9$, OR = 0.99), LCIs ($F(1,1839) = 0.484$, $p > .4$, OR = 2.40), AdvHS ($F(1,1839) = 0.461$, $p > .4$, OR = 0.76) or IntHS ($F(1,1839) = 0.002$, $p > .9$, OR = 1.02). With Polarity mood selection, there was again no effect of PrimeStatus for the SDCs ($F(1,1839) = 0.227$, $p > .6$, OR = 1.56), LCIs ($F(1,1839) = 0.000$, $p > .9$, OR = 0.98), AdvHS ($F(1,1839) = 0.002$, $p > .9$, OR = 0.99) or IntHS ($F(1,1839) = 0.603$, $p > .4$, OR = 1.31). On the basis of these results, we can be confident that PrimeStatus does not affect odds of subjunctive production for any of the groups in any of the experimental conditions. As such, we will not examine the other perspectives of the three-way interaction.

The three-way interaction between Group, MoodSelectType and PrimeStatus is represented graphically in Figure 11.

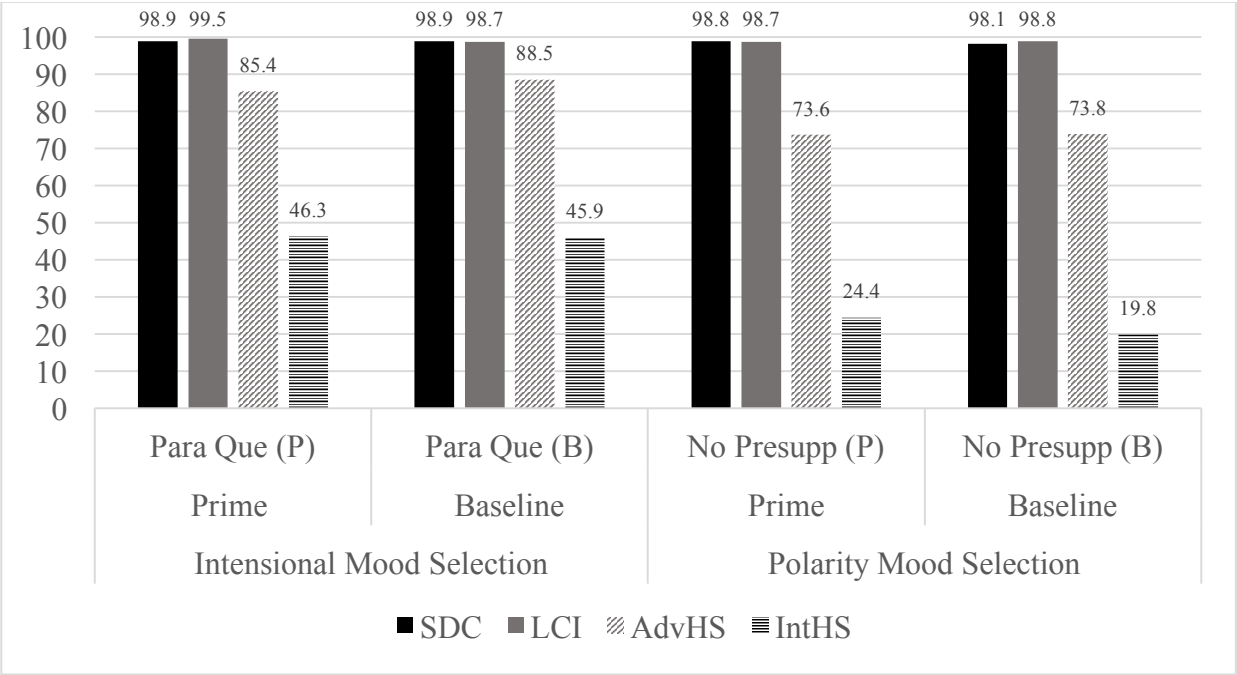

Figure 11: Probability of subjunctive: Group x MoodSelectType x PrimeStatus

Summary of three-way interaction

The three-way interaction effect, while not statistically significant, offers considerable insight into the (non)impact of PrimeStatus on participants' odds of subjunctive production. Within-group GLMMs revealed that PrimeStatus did not impact odds of subjunctive production for any of the experimental groups in any of the experimental conditions.

5.2.3.4. Summary of results from the CEPT

The results of the CEPT shed light on the three RQs of the present study. With respect to the first RQ, the results indicate that HSs' productive knowledge of subjunctive mood is variable in nature. While the Spanish-dominant controls produce subjunctive morphology nearly 100% of the time when subjunctive mood is expected, the HS groups

exhibit enormous variability in subjunctive mood production, often producing indicative mood morphology in contexts where subjunctive mood is expected . Despite this variability, statistical analysis suggests that HSs at both intermediate and advanced proficiency levels perform "above chance", demonstrating significantly higher likelihoods of producing subjunctive when subjunctive mood is expected. This mood sensitivity would not be expected if HSs' knowledge of mood is random.

The second RQ examines the role of between-group variables on HSs' subjunctive knowledge. Results of the CEPT reveal a few significant trends. As hypothesized, the SDCs (age of acquisition of English > 12) and the LCIs (age of acquisition of English between 8-12) produced significantly more subjunctive mood forms than the HS groups, suggesting that acquiring English at an earlier age increases variability in subjunctive mood production. Curiously, age effects emerged in the performance of the non-heritage groups, as the LCIs showed a unique tendency to produce subjunctive (42.2% predicted probability) in + presupposition contexts. Among the HS groups, proficiency modulated variability in subjunctive production: as hypothesized, the AdvHSs were more likely to produce subjunctive mood than the IntHSs in all subjunctive conditions.

The third RQ addressed the role of within-group variables on HSs' subjunctive mood knowledge. Results of the CEPT reveal that HSs, as hypothesized in Chapter 4, are more likely to produce subjunctive with intensional subjunctive mood than polarity subjunctive mood. Contrary to the hypotheses presented in Section 4.2, however, HSs were not more likely to produce subjunctive mood forms in the presence of structural priming. Nonetheless, HSs did exhibit a clear lexical frequency effect, producing significantly more subjunctive with frequent, as opposed to infrequent, verbs.

Up to this point, we have explored HSs' productive knowledge of subjunctive mood. In the next section, I present the results from the Contextualized Acceptability Task (CAT), which assesses HSs' receptive knowledge of subjunctive mood form.

5.3. Contextualized Acceptability Task (CAT)

Participants in the CAT rated the acceptability of subjunctive and indicative mood forms in presuppositional and non-presuppositional contexts. The results of the CAT, therefore, help illustrate HSs' *receptive* knowledge of subjunctive mood (RQ #1). By analyzing the role of between-group factors (RQ #2) and within-group factors (RQ #3), the CAT also directly addresses the second and third RQ of the present study.

5.3.1. Descriptive statistics

In the next subsections, I will achieve a few goals. I will start by identifying the independent and dependent variables in the CAT. After introducing and naming these variables, I will present the descriptive data from the experiment, which I will use to make two arguments. First, I will use the data from the Spanish-dominant controls to argue that the experimental design works as expected. Second, I will use the results from the fillers to demonstrate that all experimental groups in the study, both HSs and non-HSs, are able to attend to verbal morphology in the experimental task.

5.3.1.1. Variables in the CAT

The dependent variable in the CAT is Acceptability, measured on a 1-5 Likert scale where 1 stands for *'sounds very odd'* and 5 stands for *'sounds very good*.' The between-group independent variable is Group (SDC, LCI, AdvHS, and IntHS). The within-group

independent variables are Scenario (Found, Wrong), which refers to whether the character in the animated clip "finds" the appropriate item or not (Found = presupposition; Wrong = no presupposition), and Mood (Subjunctive, Indicative), which refers to the morphology used by the second animated character at the end of the clip. In subsequent analyses, we will also examine the within-group independent variable of Frequency (Frequent, Infrequent) to see whether, and to what extent, the frequency of the inflected verb in the adjectival relative clause affects participants' acceptance. As in Section 5.2., all variable names will be capitalized throughout the presentation of the results.

5.3.1.2. Descriptive results

All 81 participants completed all items of the CAT, which included 24 experimental items (4 conditions with 6 items each) and 16 fillers (4 conditions with 4 items each) for a total of 3240 responses. Responses were coded by the experimenter.

Before presenting the statistical models for the CAT, I will first examine the raw data in order to make two critical points. First, I will show that the SDCs performed the CAT as expected, validating the manipulations of the experiment. Second, I will argue, using the raw data from the filler conditions, that all participant groups are paying attention to manipulations in verbal morphology when rating experimental items in this task.

Recall that participants in the CAT rated sentences on a scale from 1 (*'sounds very odd'*) to 5 (*'sounds very good'*) scale. If the experiment works as planned, then the SDCs should show a clear two-way interaction between Scenario (e.g., the presence or absence of presupposition) and Mood (e.g., the mood form of the verb in the relative clause), rating indicative items much higher in the Found scenario and subjunctive items much higher in

the Wrong scenario. As shown in Table 28, the SDCs exhibit the predicted two-way interaction between Mood and Scenario, rating Indicative items higher in the Found scenario and Subjunctive items higher in the Wrong scenario.

Table 28.
Average acceptability ratings by Group, Scenario and Mood

	Found (Presupposition)		Wrong (No Presupposition)	
Group	Indicative	#Subjunctive	#Indicative	Subjunctive
SDCs	Mean: 4.45 SD: 0.56	Mean: 2.67 SD: 1.14	Mean: 2.56 SD: 0.88	Mean: 4.52 SD: 0.53
LCIs	Mean: 4.19 SD: 0.64	Mean: 3.55 SD: 0.95	Mean: 2.47 SD: 0.67	Mean: 4.32 SD: 0.74
AdvHS	Mean: 4.48 SD: 0.41	Mean: 3.52 SD: 0.74	Mean: 3.64 SD: 0.75	Mean: 4.42 SD: 0.42
IntHS	Mean: 4.39 SD: 0.55	Mean: 3.58 SD: 0.71	Mean: 3.80 SD: 0.74	Mean: 3.79 SD: 0.75

Crucially, the SDCs do not show an overall Mood bias ($M_{INDICATIVE}$ = 3.50, SD = 0.53; $M_{SUBJUNCTIVE}$ = 3.60, SD = 0.63) or an overall Scenario bias (M_{FOUND} = 3.56, SD = 0.59; M_{WRONG} = 3.54, SD = 0.54), strengthening the likelihood that they are truly judging the interaction between Mood and Scenario rather than either of these factors alone.

In the analysis that follows, I will be assuming that the HS groups, like the SDCs, are attending to verbal morphology when rating sentences in this task. To show the validity of this assumption, I will now present raw data from the filler conditions, which tested perfective and imperfective mood morphology with the verbs *querer* and *encontrar*.

If participants attend to verbal morphology when rating the filler sentences, we should expect them to (a) rate perfective items higher than imperfective items in the Found scenario and (b) rate imperfective items higher than perfective items in the Wrong scenario.

As shown in Table 29, both HS groups show this interaction, demonstrating that they are able to attend to verbal morphology when judging acceptability in the CAT items.

Table 29.
Average acceptability ratings by Group, Scenario and Aspect

Scenario	Found (Presupposition)		Wrong (No Presupposition)	
Group	Perfective *encontraste*	Imperfective *#encontrabas*	Perfective *#quisimos*	Imperfective *queríamos*
SDC	Mean: 4.95 SD: 0.17	Mean:1.39 SD: 0.58	Mean: 1.60 SD: 0.67	Mean: 4.06 SD: 0.98
LCI	Mean: 4.67 SD: 0.62	Mean: 1.49 SD: 0.52	Mean: 2.09 SD: 0.90	Mean: 3.37 SD: 0.95
AdvHS	Mean: 4.76 SD: 0.32	Mean: 1.25 SD: 0.44	Mean: 2.98 SD: 1.10	Mean: 3.68 SD: 1.15
IntHS	Mean: 4.63 SD: 0.43	Mean: 2.23 SD: 1.43	Mean: 3.31 SD: 1.13	Mean: 3.66 SD: 1.20

5.3.2. Statistical models for the CAT

As with the CEPT, I will use GLMMs to analyze the results of the CAT. GLMMs were used instead of Mixed ANOVAs for a number of different reasons, primarily due to the non-continuous nature of the dependent variable.

In the subsections that follow, I will start by describing how I transformed the scalar acceptability data into a format appropriate for use with GLMMs. Then, I will present GLMMs which test the role of the independent variables presented in Section 5.3.1.1 on participants' likelihood of accepting experimental sentences in the CAT.

5.3.2.1. Transforming the raw data

Participants' scalar responses were converted to a binary scale scale such that all responses were coded as either 0 ('reject') or 1 ('accept'). Ultimately, I chose to reclassify ratings of 1 ('sounds very odd'), 2 ('sounds odd') and 3 ('sounds okay') as 0 ('reject') and ratings of 4 ('sounds good') and 5 ('sounds very good') as 1 ('accept'). I chose to label 3 as an instance of 'reject' for two main reasons. First of all, excluding participants' 3 ratings

would lead to the loss of 206 responses (10.6% of the CAT data). Second, an analysis of participant responses revealed that HSs used the 3 rating more often in sentences which are expected to be infelicitous, suggesting that they are using this middle point of the scale as a marker of rejection[28]. Supporting this claim is the fact that 67.5% of the 3 ratings provided by the HS groups were given in infelicitous conditions.

5.3.3. Results from the GLMMs

In Sections 5.3.3.1 through 5.3.3.4, I will present the GLMMs which were used to evaluate the results of the CAT. The title of each subsection will indicate the independent variables, presented in all caps, as well as any variables that are excluded from a particular analysis. In the body of each subsection I will discuss main effects, as well as two- and three-way interaction effects. To do so, I will present participants' predicted acceptability data, providing tables with log-odds, standard errors, and odds, as well as predicted probabilities and 95% confidence intervals. At the end of each model, I will supplement the statistical analyses by providing a graph that illustrates the effects of each of the independent variables on participants' acceptability patterns.

5.3.3.1. Group, Mood, and Scenario

The data were analyzed with a GLMM (GLMM #1), using the GENLINMIXED procedure in SPSS 24. The independent variables in GLMM #1 were Group (AdvHS, IntHS, SDC, LCI), Scenario (Found, Wrong) and Mood (Subjunctive, Indicative) and the binary dependent variable was Acceptance (Accept or Reject). The model also included

[28] See Polinsky (2016) for a discussion of HSs' tendency to use the upper end of acceptability scales.

random subject and item intercepts, which allow the researcher to control for random variance among subjects (i.e., some subjects are more accepting in general) and lexical items (i.e., some verbs are more accepted by participants, regardless of condition).

5.3.3.1.1. Main effects

The GLMM revealed significant main effects of Group ($F(3,76) = 3.40$, $p < .05$) and Scenario ($F(1,1928) = 7.953$, $p < .01$) but not Mood ($F(1,1928) = .04$, $p > .8$. Pairwise comparisons for Group, shown in Table 30, revealed that, in general, AdvHSs' odds of accepting a given item in the CAT were significantly higher than the odds of the SDC's ($p < .05$; OR = 1.98) and the LCIs ($p < .01$, OR = 2.71). Overall, AdvHSs' odds of accepting experimental items were about 2 times higher than the odds of the SDCs and about 2.7 times higher than the odds of the LCIs. All other group comparisons were non-significant.

Table 30.
Probability of acceptance by Group

Group	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
SDC	0.690	0.326	1.994	66.6%	50.2%	79.8%
LCI	0.380	0.322	1.462	59.4%	42.6%	74.2%
AdvHS	1.375	0.318	3.955	79.8%	66.9%	88.5%
IntHS	1.007	0.320	2.737	73.3%	58.2%	84.3%

Pairwise comparisons for Scenario indicated that participants' odds of accepting a given experimental item were higher in the Found scenario than in the Wrong scenario, $p < .01$, OR = 1.42, as shown in Table 31. The OR of 1.42 means that participants were about 1.5 times more likely to accept Found scenario items than Wrong scenario items.

Table 31.
Probability of acceptance by Scenario

Scenario	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
Found	1.037	.259	2.821	73.8%	60.7%	83.7%
Wrong	0.689	.260	1.992	66.6%	52.2%	78.4%

Finally, the lack of a significant main effect for Mood ($p > .8$, OR = .98) indicates that participants' odds of accepting a given item were not higher for Subjunctive or Indicative mood items, respectively. This lack of a main effect, shown in Table 32, is exactly what is expected in a task where half of the conditions favor the acceptance of subjunctive item and half favor the acceptance of indicative.

Table 32.
Probability of acceptance by Mood

Mood	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
Indicative	0.851	0.260	2.342	70.1%	56.2%	81.0%
Subjunctive	0.875	0.259	2.399	70.6%	56.8%	81.3%

Summary of main effects

The AdvHS had higher odds of accepting experimental items in the CAT than the SDCs or LCIs, demonstrating a slight acceptability bias. Additionally, participants in the CAT showed significantly higher odds of acceptance for Found scenario items, raising the possibility that participants misunderstood the instructions and accepted or rejected items based on Scenario (e.g., whether Nico brought back the right item), rather than the interaction between Scenario and Mood. There was no main effect of Mood, however, suggesting that participants' odds of accepting an item were not significantly higher with Indicative mood than Subjunctive mood, respectively.

5.3.3.1.2. Two-way interactions

The GLMM also revealed significant interaction effects for all two way interactions including Group x Scenario ($F(3,1928)=3.735$, $p < .05$), Group x Mood ($F(3,1928) = 7.865$, $p < .001$), and Scenario x Mood ($F(1,1928) = 288.269$, $p < .001$. For all interactions,

separate GLMMs were run on the differences in odds across the relevant conditions to make relevant between- and within-group pairwise comparisons.

We start our discussion by examining the Group x Scenario interaction (Table 33), which sheds light on whether participants in each of the groups exhibit higher likelihood of accepting items in either the Found or Wrong scenarios, respectively. The results of a separate GLMM revealed significant effects of Scenario for the IntHS ($F(1,1928) = 5.077$, $p < .05$; OR = 1.68) and the LCIs ($F(1,1928) = 16.255$, $p < .001$; OR = 2.55), but not the AdvHSs ($p > .7$, OR = 0.94) or the SDCs ($p > .9$, OR = 1.00). In other words, both the IntHSs and the LCIs had significantly higher odds of accepting items in the Found scenario as opposed to the Wrong scenario. The IntHSs were about 1.7 times more likely to accept items in the Found scenario while the LCIs were 2.5 times more likely to accept items in the Found scenario. It appears, therefore, that both the LCIs and IntHSs show a Scenario bias, showing significantly higher odds of accepting Found condition items.

Table 33.
Probability of acceptance: Group x Scenario

Group	Scenario	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
SDC	Found	0.689	0.347	1.992	66.6%	49.3%	80.3%
	Wrong	0.691	0.358	1.996	66.6%	48.9%	80.6%
LCI	Found	0.849	0.342	2.337	70.0%	53.5%	82.6%
	Wrong	-0.089	0.343	0.915	47.8%	31.0%	65.1%
AdvHS	Found	1.343	0.342	3.831	79.3%	65.4%	88.6%
	Wrong	1.407	0.344	4.084	80.3%	66.7%	89.3%
IntHS	Found	1.266	0.346	3.547	78.0%	63.4%	87.9%
	Wrong	0.748	0.333	2.113	67.9%	51.3%	80.9%

To explore whether participant groups showed a Mood bias, we next consider the Mood x Group interaction (Table 34). This interaction effect ignores the role of Scenario and considers the possibility that some groups, in general, show a higher acceptance of experimental items with Subjunctive or Indicative mood, respectively.

Results of a separate GLMM revealed a significant effect of Mood for the IntHS ($F(1,1928) = 15.128$, $p < .001$, OR = 2.46) and the LCIs ($F(1,1928) = 6.459$, $p < .05$, OR = 0.55) but not the SDCs ($p > .1$, OR = 0.69) or the AdvHS ($p > .9$, OR = 0.97). While both the IntHS and the LCIs exhibited mood bias, the *direction* of that bias differed across groups. The IntHS' OR of 2.46 indicates that they were approximately 2.5 times more likely to accept Indicative items as opposed to Subjunctive items. The LCIs, on the other hand, were about half (0.55) as likely to accept Indicative items.

These results help us to explain the similar behavior of the IntHS and the LCIs in the Group by Scenario interaction discussed above. Both the IntHSs and the LCIs are more likely to accept Found items but for quite different reasons. The IntHS' bias towards accepting more items in the Found conditions stems from their *decreased* likelihood of accepting target-like subjunctive items in the Wrong scenario. (By rejecting Wrong subjunctive items, their *relative* acceptance of Found scenario items increased.) The LCIs' Found item bias, on the other hand, results from their *increased* likelihood of accepting non-target subjunctive items in the Found scenario. (By accepting Found subjunctive items, their relative acceptance of Found scenario ratings increased.)

Table 34.
Probability of acceptance: Group x Mood

Group	Mood	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
SDC	Indic.	0.501	0.349	1.650	62.3%	44.5%	77.2%
	Subj.	0.879	0.356	2.408	70.7%	53.7%	83.4%
LCI	Indic	0.085	0.346	1.089	52.1%	34.7%	69.0%
	Subj	0.675	0.338	1.964	66.3%	49.3%	79.9%
AdvHS	Indic	1.360	0.342	3.896	79.6%	65.7%	88.8%
	Subj	1.390	0.344	4.015	80.0%	66.3%	89.1%
IntHS	Indic	1.456	0.348	4.289	81.1%	67.7%	89.8%
	Subj	0.558	0.332	1.747	63.6%	46.6%	77.7%

The last two-way interaction, shown in Table 35, reveals whether participants, as a whole, are accepting items on the basis of an interaction between Scenario and Mood. Essentially, this interaction tells us whether participants are more or less likely to accept Subjunctive and Indicative mood items based on Scenario. For the Found scenario, a separate GLMM revealed a highly significant effect of Mood, $F(1,1928)=147.994$, $p < .001$, OR = 8.57. Pairwise comparisons revealed that in the Found scenario, participants were about 8.5 times more likely to accept indicative, rather than subjunctive mood items. A separate GLMM for the Wrong scenario also revealed a highly significant effect of Mood, $F(1,1928) = 151.625$, $p < .001$, OR = 9.00. This time, the pairwise comparison revealed that participants were about 9 times more likely to accept subjunctive, as opposed to indicative mood items, in the Wrong scenario.

Table 35.
Probability of acceptance: Scenario x Mood

Scenario	Mood	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
Found	Indic	2.111	0.284	8.256	89.2%	81.6%	93.9%
	Subj	-0.038	0.264	0.963	49.0%	34.5%	63.8%
Wrong	Indic	-0.410	0.267	0.664	39.9%	26.6%	54.9%
	Subj	1.788	0.282	5.977	85.7%	76.2%	91.7%

This interaction effect gives further reason to believe in the validity of the experimental design. Clearly, participants are differentially accepting subjunctive and indicative mood items based on the presuppositional status of the Scenario. Nonetheless, it is important to remember that this two-way interaction does not take into account the role of Group. Only by examining the three-way interaction between Mood, Scenario and Group can we determine whether the groups differ in their acceptance of mood morphology across Found and Wrong scenarios, respectively.

Summary of two-way interactions

Each of the three two-way interactions was statistically significant. Further analysis of the Group by Scenario interaction revealed that only the IntHS and the LCIs showed a Scenario bias, e.g., higher odds of accepting items in the Found scenario. Underlying these two groups' biases, however, were two completely different tendencies. An analysis of the Group by Mood interaction revealed that the IntHS showed significantly higher odds of accepting Indicative, as opposed to Subjunctive, items. The LCIs, on the other hand, had significantly higher odds of accepting Subjunctive items. Finally, the Scenario by Mood interaction showed that participants' odds of accepting Subjunctive and Indicative items varied by Scenario: in the Found scenario, participants' odds of accepting an item were significantly higher with Indicative mood while in the Wrong scenario, their odds of accepting an item were significantly higher with Subjunctive mood.

5.3.3.1.3. Three-way interaction

Finally, the GLMM revealed a significant effect of Group x Scenario x Mood, $F(3, 1928) = 16.334$, $p < .001$. Separate GLMMs were run to determine whether each of the groups showed a higher probability of accepting an item based on the interaction between its Mood and Scenario. We first consider whether the groups show mood sensitivity in the Found scenario, where they're expected to be more accepting of Indicative mood items.

In the Found scenario, all groups were significantly more likely to accept indicative, as opposed to subjunctive items: SDCs: $F(1,1928) = 71.069$, $p < .001$, OR = 22.18; LCIs: $F(1,1928) = 20.960$, $p < .001$, OR = 4.45; AdvHS: $F(1,1928) = 36.609$, $p < .001$, OR = 8.70; IntHS: $F(1, 1928) = 27.143$, $p < .001$, OR = 6.30. While all groups are more likely to accept indicative items in this scenario, the SDCs, who are 22 times more likely to accept

indicative rather than subjunctive items, make this distinction most clearly. Notably, both the AdvHS (OR = 8.70) and the IntHS (OR = 6.30) make a stronger distinction in this scenario than the LCIs (OR = 4.45), largely due to the LCIs' increased likelihood of accepting subjunctive mood items in the Found scenario.

If we narrow our focus, we can examine whether the groups differed from one another in their odds of accepting items in the FoundIndicative and FoundSubjunctive conditions. A separate GLMM revealed that the four groups did not differ in their odds of accepting FoundIndicative items, $F(3,368) = 1.16$, $p > .3$. With FoundSubjunctive items, however, a GLMM revealed significant effects of Group, $F(3,165) = 3.844$, $p < .05$. Pairwise comparisons showed that the AdvHS ($p < .01$, OR = 3.07), the IntHS ($p < .01$, OR = 3.34) and the LCIs ($p < .05$, OR = 2.62) were all significantly more likely to accept subjunctive mood items in the Found scenario than the SDCs were. Though the HSs pattern with the LCIs here, their similar behavior may be driven by different underlying causes. The HSs' increased acceptance of these items could be the result of their general acceptability bias. The LCIs' higher acceptance, however, may be driven by a more systematic underlying acceptance of subjunctive mood forms in the context of presupposition[29]. This possibility is supported by the data from the CEPT, which showed that the LCIs produced subjunctive nearly half the time in presuppositional contexts.

We now turn towards the Wrong Scenario, where participants are expected to show higher acceptance of Subjunctive items. In the Wrong scenario, the SDCs ($F(1,1928) = 88.182$, $p < .001$, OR = 47.27), the LCIs ($F(1,1928) = 62.183$, $p < .001$, OR = 14.44) and

[29] Note that the LCIs' higher acceptance of FoundSubjunctive items, relative to the SDCs, is not attributable to a higher overall tendency towards acceptance: The LCIs' overall predicted probability of acceptance for items in the CAT (59.4%) is 7.2 points lower than the overall probability acceptance of the SDCs (66.6%).

the AdvHS (F(1,1928) = 37.029, $p < .001$, OR = 9.23) were significantly more likely to accept subjunctive, as opposed to indicative mood items. The IntHS, however, did not show a higher likelihood to accept subjunctive (vs. indicative) items, F(1,1928) = .022, $p > .8$, OR = 1.04, revealing non-target-like behavior in this scenario. Of the three groups that show a higher acceptance of subjunctive items here, the SDC's do so most clearly. Their OR of 47.27 is over three times higher than the OR for the LCIs (OR = 14.44) and over five times higher than the OR for the AdvHS (OR = 9.23). Unlike in the Found scenario, the LCIs' perform more closely to the SDCs than the AdvHS do.

If we focus our attention on each of the Mood forms in the Wrong condition, we can see that the groups differed in their odds of accepting items with (a) indicative mood (F(3,189) = 13.552, $p < .001$) and (b) subjunctive mood (F(3,311) = 7.603, $p < .001$.) Starting with Indicative mood, pairwise comparisons revealed that the AdvHS (vs. SDCs: $p < .001$, OR = 4.63; vs. LCIs: $p < .001$, OR = 5.58) and the IntHS (vs. SDCs: $p < .001$, OR = 7.13; vs. LCIs: $p < .001$, OR = 8.58) were significantly more likely than the SDCs and the LCIs to accept infelicitous indicative items in the Wrong scenario. This finding is consistent with the claim that for some HSs, indicative mood is a default mood form.

Additional pairwise comparisons revealed Group differences in the WrongSubjunctive items as well. The AdvHS were more likely to accept these items than the IntHS ($p < .001$, OR = 5.74) and the LCIs ($p < .01$, OR = 3.56) but not the SDCs ($p > .05$; OR = 1.11). The fact that the AdvHS did not differ from the SDCs in their odds of accepting subjunctive items in the Wrong scenario is noteworthy but likely attributable to the fact that the items in this condition are expected to be felicitous and accepted. The IntHS, on the other hand, were less likely to accept these items than the SDC's ($p < .001$,

OR = 0.16) but not the LCIs ($p > .20$, OR = .62), revealing a considerable default preference for indicative mood forms. Finally, the SDCs were more likely than the LCIs to accept WrongSubjunctive items, $p < .01$, OR = 3.94, a somewhat surprising finding that may be rooted in the LCIs' tendency to reject more items, in general, than the SDCs.

Table 36.
Probability of acceptance: Group x Scenario x Mood

Group	Scenario	Mood	Logodds	SE	Odds	Probability	CI-Lower	CI-Upper
SDC	Found	indic	2.239	.419	9.384	90.4%	80.2%	95.6%
		subj	-.861	.365	.423	29.7%	16.7%	47.2%
	Wrong	indic	-1.237	.374	.290	22.5%	11.9%	38.3%
		subj	2.618	.448	13.708	93.2%	84.8%	97.1%
LCI	Found	indic	1.595	.393	4.928	83.1%	68.9%	91.6%
		subj	.102	.363	1.107	52.6%	34.5%	70.0%
	Wrong	indic	-1.424	.384	.241	19.4%	10.0%	34.4%
		subj	1.247	.381	3.480	77.7%	61.6%	88.3%
AdvHS	Found	indic	2.424	.421	11.291	91.9%	82.9%	96.3%
		subj	.261	.345	1.298	56.5%	38.9%	72.6%
	Wrong	indic	.296	.346	1.344	57.4%	39.7%	73.3%
		subj	2.518	.429	12.404	92.5%	84.0%	96.7%
IntHS	Found	indic	2.186	.415	8.900	89.9%	79.4%	95.4%
		subj	.346	.360	1.413	58.6%	40.3%	74.7%
	Wrong	indic	.726	.364	2.067	67.4%	49.5%	81.3%
		subj	.770	.365	2.160	68.3%	50.6%	82.0%

The results of the three-way interaction between Group, Scenario and Mood are presented in Table 36, as well as in Figure 12 below.

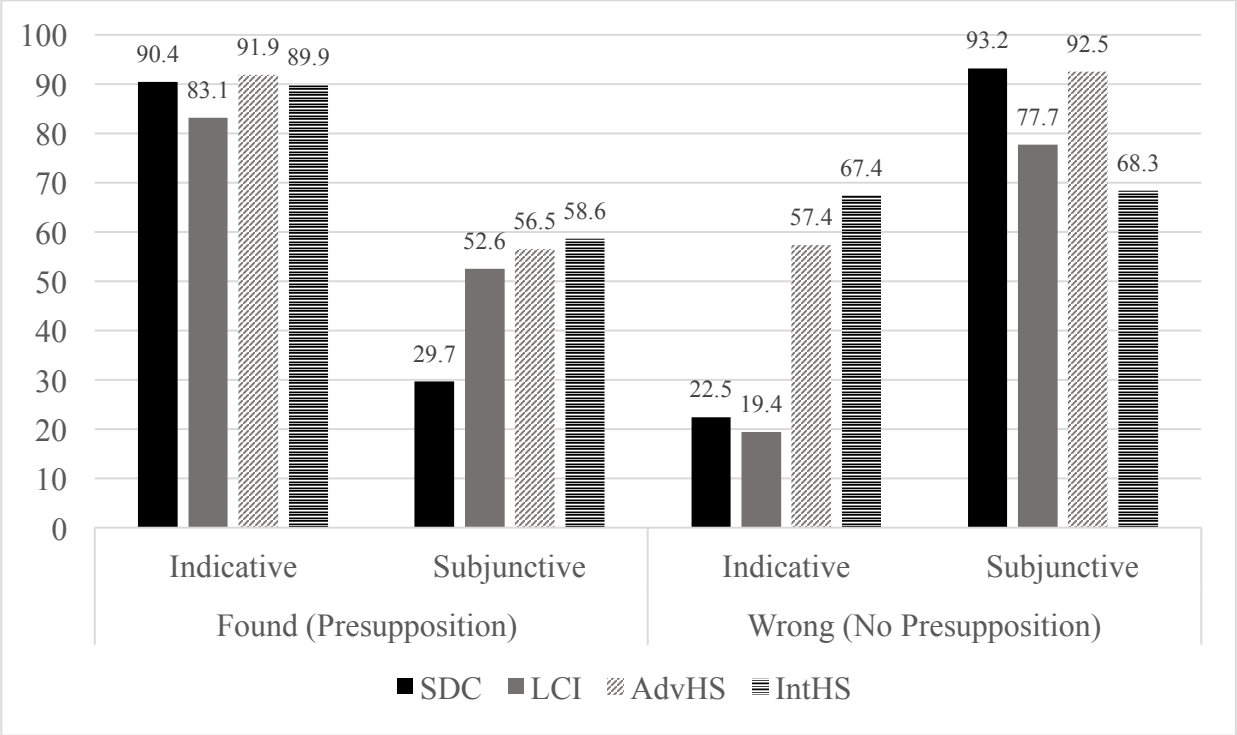

Figure 12: Probability of acceptance: Group x Scenario x Mood

Summary of three-way interaction

In the Found scenario, participants in all groups showed significantly higher odds of accepting Indicative, as opposed to Subjunctive, mood items. Nonetheless, the LCIs, AdvHS and IntHS were all significantly more likely to accept non-target Subjunctive items in this condition than the SDCs. In the Wrong scenario items, on the other hand, only the SDCs, LCIs, and AdvHS had significantly higher odds of accepting Subjunctive mood items. Even though the AdvHS demonstrated higher odds of accepting Subjunctive items in the Wrong scenario, both the AdvHSs' and the IntHSs' odds of accepting infelicitious WrongIndicative items were significantly higher than the odds for both of the Spanish-dominant groups, revealing indicative mood bias.

5.3.3.2. Group, Frequency and Mood (Found)

GLMM #1 revealed participants' sensitivity to the Mood x Scenario interaction. GLMM #2 and GLMM #3 build on this finding by whether this sensitivity to mood morphology is conditioned by the frequency of the inflected verb forms. Are participants more sensitive to mood distinctions with frequent or infrequent verb forms?

The dependent variable, as in GLMM #1, is Acceptability (Accept or Reject). The independent variables are Group (SDCs, LCIs, AdvHSs, or IntHSs), Mood (Indicative or Subjunctive), and Frequency (Frequent or Infrequent). Frequency status for each experimental item was determined by consulting Davies' (2006) frequency rankings and categorizing each of the six experimental verbs into two categories: Frequent, which included *ser* (Davies' ranking: 8), *estar* (Davies' ranking: 17), and *tener* (Davies' ranking: 18), and Infrequent, which included *venir* (Davies' ranking: 105), *costar* (Davies' ranking: 626), and *funcionar* (Davies' ranking: 692). The average Davies frequency ranking of the Frequent items was 14.33 while the average frequency ranking of the Infrequent items was 474.33. In order to add Frequency as a variable in the model, two separate GLMMs were run. Consequently, the next GLMM (GLMM #2) focuses only on the Found scenario items while the final GLMM (GLMM #3) focuses on the Wrong scenario.

To avoid repeating the findings from GLMM #1, I will focus exclusively on main and interaction effects involving the newly introduced variable Frequency.

5.3.3.2.1. Main effects

GLMM #2 revealed a non-significant main effect of Frequency, $F(1,4) = 3.486$, $p > .1$. Participants' odds of accepting items in the Found scenario were not conditioned by

Frequency. Because the main effect does not take into account Mood or Group, we now turn to explore two-way interactions between Frequency and each of these variables.

5.3.3.2.2. Two-way interactions

GLMM #2 revealed a statistically significant two-way interaction of Mood and Frequency, $F(1,956) = 5.524$, $p < .05$. With Indicative mood items, there was no significant effect of Frequency, $F(1,7) = .552$, $p > .4$, OR =1.41), meaning that participants' odds of accepting a given item were not significantly higher with Frequent or Infrequent verbs, respectively. With Subjunctive mood items, however, there was a statistically significant effect of Frequency, $F(1,5) = 7.756$, $p < .05$, OR = 3.22. Participants' odds of accepting a given item were over three times higher with Infrequent, as opposed to Frequent, verbs.

The Group x Frequency interaction, which illuminates whether the variable of Frequency operates differently across the different participant groups, was not statistically significant, $F(3,956) = .944$, $p > .9$, meaning that the four participant groups' odds of acceptance are only slightly, though not significantly, higher with Infrequent verbs than with Frequent verbs (all OR's between 1.93 and 2.30).

Summary of two-way interactions

The statistically significant Mood by Frequency interaction revealed that Frequency significantly impacts participants' odds of acceptance, though only with Subjunctive mood items. The Group by Frequency interaction, on the other hand, shows that Frequency does not have a differential impact on the four participant groups, all of whom accept items with Infrequent verbs slightly more than items with Frequent verbs.

5.3.3.2.3. Three-way interaction

The three-way interaction Group x VerbFrequency x Mood was not statistically significant, F(3,956) = .060, $p > .9$. Nonetheless, additional GLMMs were run to make a different comparisons about Frequency and its impact on participants' demonstrated sensitivity to mood. The remainder of Section 5.3.3.2 will proceed as follows.

First, we will look at within-group comparisons, determining whether Frequency impacts each group's odds of acceptance with both Indicative and Subjunctive mood items. Second, we will take a between-group perspective, examining whether and to what extent the experimental groups differ from one another in each of the four conditions: IndicativeFrequent, IndicativeInfrequent, SubjunctiveFrequent and SubjunctiveInfrequent.

Table 37.
Probability of acceptance: Group x Frequency x Mood (Found)

Group	Frequency	Mood	Logodds	SE	Odds	Probability	CI-Low	CI-Upper
SDC	High	indic	2.091	.523	8.093	89.0%	73.8%	95.9%
		subj	-1.455	.475	.233	18.9%	8.1%	38.1%
	Low	indic	2.248	.539	9.469	90.4%	76.2%	96.6%
		subj	-0.302	.438	.739	42.5%	22.9%	64.7%
LCI	High	indic	1.436	.485	4.204	80.8%	61.0%	91.9%
		subj	-0.473	.451	.623	38.4%	19.7%	61.2%
	Low	indic	1.807	.509	6.092	85.9%	68.5%	94.5%
		subj	0.725	.457	2.065	67.4%	44.6%	84.1%
AdvHS	High	indic	2.176	.515	8.811	89.8%	75.7%	96.1%
		subj	-0.277	.423	.758	43.1%	23.8%	64.8%
	Low	indic	2.717	.583	15.135	93.8%	82.5%	98.0%
		subj	0.846	.433	2.330	70.0%	48.6%	85.2%
IntHS	High	indic	1.956	.510	7.071	87.6%	71.6%	95.2%
		subj	-0.234	.438	.791	44.2%	24.2%	66.3%
	Low	indic	2.262	.540	9.602	90.6%	76.4%	96.6%
		subj	.969	.453	2.635	72.5%	50.9%	87.0%

We begin our discussion of the three-way interaction, shown in Table 37, by zeroing in on the Indicative mood items, where Frequency did not have a statistically

significant impact for the SDCs ($F(1,33) = .053$, $p > .8$, OR = 1.17), LCIs ($F(1,23) = .357$, $p > .5$, OR = 1.45), AdvHS ($F(1,41) = .572$, $p > .4$, OR = 1.72) or IntHS ($F(1,31) = .208$, $p > .6$, OR = 1.36). While none of the groups' odds of acceptance were significantly affected by Frequency, all groups show a slightly higher probability (OR's between 1.17 and 1.72) of accepting items with Infrequent verbs.

With Subjunctive mood items, however, there was a statistically significant effect of Frequency on odds of acceptance for the LCIs ($F(1,15) = 4.577$, $p < .05$, OR = 3.31) and the IntHS ($F(1,14) = 4.770$, $p < .05$, OR = 3.33), both of whom were over three times more likely to accept Infrequent items than Frequent items. While the effect of Frequency was only marginally significant for the SDCs ($F(1,16) = 4.123$, $p < .06$, OR = 3.17) and the AdvHS ($F(1,12) = 4.488$, $p < .06$, OR = 3.07), both of these groups were also over three times more likely to accept items with Infrequent verbs.

Thus far, we have focused on the within-group perspective of the three-way interaction, showing that Frequency has a strong impact on acceptance for all of the experimental groups, though only with Subjunctive mood items. We now turn to the between-group perspective of this interaction, which considers the extent to which the groups differ from one another in each of the four combinations of Mood by Frequency.

There were no significant differences between the groups' odds of acceptance with IndicativeFrequent ($F(3,411) = .621$, $p > .6$) or IndicativeInfrequent ($F(3,560) = .601$, $p > .6$) forms, all p's > .18 and all OR's between 1.01 and 2.5, showing that the groups are similarly accepting of Indicative items with both Frequent and Infrequent verb forms.

With Subjunctive items, however, group differences emerged. With SubjunctiveFrequent items, there was a marginally significant effect of Group, $F(3,210) =$

2.220, $p < .08$. Pairwise comparisons revealed that the AdvHSs' ($p < .05$, OR = 3.25) and IntHSs' ($p < .05$, OR = 3.39) odds of accepting SubjunctiveFrequent items were over three times higher than the same odds for the SDCs. While there were no statistically significant differences between the SDCs and LCIs in this condition ($p < .08$, OR = 2.67), the LCIs were nearly three times more likely to accept SubjunctiveFrequent items than the SDCs. This finding, in conjunction with the results of the CEPT, suggests that subjunctive mood morphology is acceptable for the LCIs even in contexts with clear presupposition.

With SubjunctiveInfrequent items, an even stronger effect of Group emerged, $F(3, 207) = 2.670$, $p < .05$. Pairwise comparisons revealed that the LCIs, ($p < .05$, OR = 2.79), AdvHS ($p < .05$, OR = 3.15), and IntHS ($p < .05$, OR = 3.56) all were approximately three times more likely to accept these items than the SDCs. It is noteworthy that the LCIs, who have thus far shown a marked tendency to overproduce (CEPT) and overaccept (CAT) subjunctive forms in Found (presupposition) contexts, only differ significantly from the SDCs in the SubjunctiveInfrequent items.

Figure 13 presents the results of the three-way interaction between Group, Mood and Frequency in GLMM #2 in graphical form. Recall that this interaction focuses on the Found scenario items only.

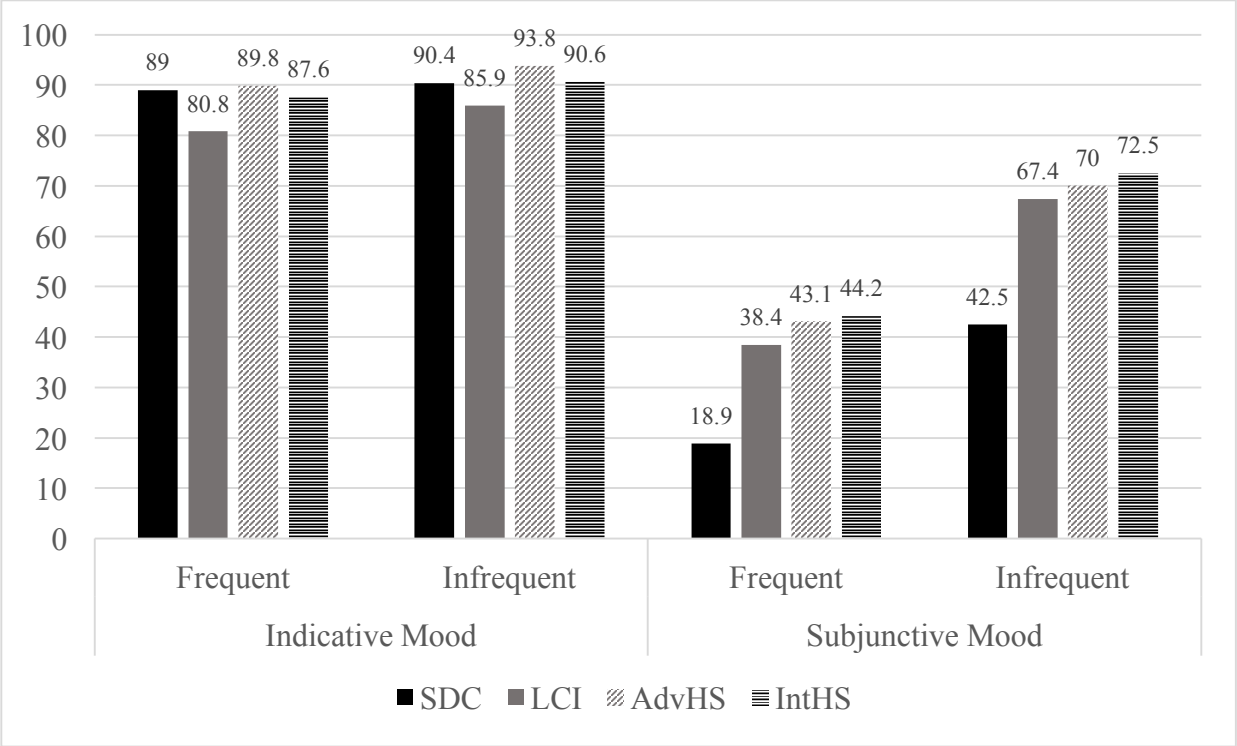

Figure 13: Probability of acceptance: Group x Frequency x Mood (Found)

Summary of three-way interaction

Analyses of the three-way interaction revealed the following trends. First, Frequency did not significantly impact any of groups' odds of acceptance with Indicative mood items. With Subjunctive mood items, however, all groups were at least three times more likely to accept items with Infrequent verbs, providing robust evidence that Frequency affects the acceptance patterns of all experimental groups. Between-group comparisons revealed no differences across the groups with Indicative mood items, regardless of Frequency. With Subjunctive mood items, however, considerable differences emerged. With Frequent verbs, the HSs' odds of acceptance were significantly higher than the odds for the SDCs. With Infrequent verbs, however, the HSs' *and* LCIs' odds of acceptance were significantly higher than the odds for the SDCs.

5.3.3.3. Group, Frequency and Mood (Wrong)

GLMM #2 examined the effect of Frequency on participants' acceptance of experimental items in the Found scenario. GLMM #3 complements GLMM #2 by testing the effect of Frequency on participants' acceptance of items in the Wrong scenario.

The dependent variable, as in GLMMs #1 and #2, is Acceptability (Accept or Reject). The independent variables are Group (SDCs, LCIs, AdvHSs, or IntHSs), Mood (Indicative or Subjunctive), and Frequency (Frequent or Infrequent). As with GLMM #2, Frequency status was determined by consulting Davies' (2006) frequency rankings and categorizing each of the six experimental verbs into two categories: Frequent, which included *ser*, *estar*, and *tener*, and Infrequent, which included *venir*, *costar*, and *funcionar*.

To avoid repeating the findings from GLMM #1 I will focus exclusively on main and interaction effects involving the independent variable, Frequency.

5.3.3.3.1. Main effects

GLMM #3 revealed a non-significant main effect of Frequency, ($F(1,4) = 1.445$, $p > .2$), meaning that participants' odds of acceptance were not significantly different in items with Frequent and Infrequent verbs, respectively.

5.3.3.3.2. Two-way interactions

While there was no significant main effect of Frequency, we now explore the two-way interactions involving Frequency to determine if Frequency has a differential impact on (a) some groups more than other groups and/or (b) one mood form more than another.

GLMM #3 revealed a non-significant interaction between Group and Frequency (F (3,956) = 0.510, $p > .6$), indicating that the participant groups did not differ from one another in their slightly, though not statistically-significant, higher odds of acceptance for items with Infrequent verbs. The model also revealed a non-significant interaction between Mood and Frequency, F (1,956) = 0.237, $p > .6$). With both Indicative items (F (1,5) = 1.794, $p > .2$) and Subjunctive items (F (1,6) = 0.883, $p > .6$), Frequency did not significantly affect participants' odds of acceptance.

Summary of two-way interactions

Neither of the two-way interactions involving Frequency were statistically significant. This suggests that however Frequency affects odds of acceptance, this effect is not statistically different across groups or across mood forms. To get a complete picture of the effect of Frequency, however, we must examine the three-way interaction.

5.3.3.3.3. Three-way interaction

The three-way interaction Group x Frequency x Mood (Table 38) was not statistically significant, F(3,956) = .296, $p > .8$. Nonetheless, additional GLMMs were run to evaluate the role of Frequency and its impact on participants' demonstrated sensitivity to mood. The remainder of this section will proceed as follows.

First, we will look at within-group comparisons, determining whether Frequency impacts each group's odds of acceptance with both Indicative and Subjunctive mood items. Second, we will take a between-group perspective, examining whether and to what extent the experimental groups differ from one another in each of the four conditions: IndicativeFrequent, IndicativeInfrequent, SubjunctiveFrequent and SubjunctiveInfrequent.

Table 38.
Probability of acceptance: Group x Frequency x Mood (Wrong)

Group	Frequency	Mood	Logodds	SE	Odds	Probability	CI Low	CI Up
SDC	High	indic	-1.451	.505	.234	19.0%	7.4%	40.6%
		subj	2.259	.568	9.574	90.5%	74.8%	96.9%
	Low	indic	-.909	.481	.403	28.7%	12.5%	53.2%
		subj	2.784	.649	16.184	94.2%	81.4%	98.4%
LCI	High	indic	-1.745	.529	.175	14.9%	5.5%	34.6%
		subj	1.169	.501	3.219	76.3%	52.6%	90.3%
	Low	indic	-1.024	.490	.359	26.4%	11.2%	50.7%
		subj	1.215	.499	3.370	77.1%	53.8%	90.7%
AdvHS	High	indic	-.147	.458	.863	46.3%	23.9%	70.4%
		subj	2.035	.532	7.652	88.4%	71.6%	95.9%
	Low	indic	.734	.463	2.083	67.6%	43.0%	85.2%
		subj	3.158	.706	23.524	95.9%	85.2%	99.0%
IntHS	High	indic	.373	.476	1.452	59.2%	34.1%	80.3%
		subj	.538	.478	1.713	63.1%	37.8%	82.8%
	Low	indic	1.027	.488	2.793	73.6%	49.5%	88.8%
		subj	.936	.485	2.550	71.8%	47.3%	87.9%

We begin our analysis of the three-way interaction by focusing our attention on Indicative mood items. GLMM #3 revealed that with Indicative mood items, Frequency does not significantly affect odds of acceptance for the SDCs ($F(1,11) = .696$, $p > .4$, OR = 1.72), LCIs ($F(1,13) = 1.150$, $p > .3$, OR = 2.06), AdvHS ($F(1,8) = 2.112$, $p > .1$, OR = 2.41) or IntHS ($F(1,10) = 1.068$, $p > .3$, OR = 1.92). The non-significant *p*-value for this interaction shows that with Indicative items, whatever effect Frequency has on acceptance is not significantly different for the different groups.

With Subjunctive mood items, the same pattern holds. GLMM #3 revealed a non-significant effect of Frequency on odds of acceptance for the SDCs ($F(1,28) = .408$, $p > .5$, OR = 1.69), LCIs ($F(1,12) = .005$, $p > .9$, OR = 1.05), AdvHS ($F(1,32) = 1.752$, $p > .1$, OR = 3.07), and IntHS ($F(1,10) = .396$, $p > .5$, OR = 1.49). Interestingly, the AdvHSs' odds of acceptance were three times higher (OR = 3.07) for items with Infrequent verbs, perhaps

because they were unable to recognize the mood morphology on these less frequent items. (In other words, they may accept an infrequent form like *funcione* without necessarily knowing that it is a subjunctive form.)

Thus far, we have focused on the within-group perspective of the three-way interaction, showing that Frequency does not significantly impact acceptance for any of the experimental groups. We now examine the between-group perspective of this interaction, to see if the groups differ from each other in each of the four Mood x Frequency conditions.

In the Indicative mood items, there was a statistically significant effect of Group on items with both Frequent ($F(3,346) = 7.785$, $p < .001$) and Infrequent ($F(3,301) = 9.660$, $p < .001$) verbs, respectively. In the IndicativeFrequent condition, pairwise comparisons revealed that the AdvHS' (vs. SDC: $p < .01$, OR = 3.68; vs. LCI: $p < .01$, OR = 4.94) and IntHS' (vs. SDC: $p < .001$, OR = 6.20; vs. LCI: $p < .001$, OR = 8.31) odds of acceptance were significantly higher than the odds for the Spanish-dominant groups. The IntHS' odds of accepting these items were over 1.6 times higher than the odds for the AdvHS, but this difference was not statistically significant, $p > .2$. In the IndicativeInfrequent condition, a similar pattern emerges, as the AdvHSs' (vs. SDCs: $p < .01$, OR = 5.17; vs. LCIs: $p < .001$, OR = 5.80) and the IntHS' (vs. SDCs: $p < .001$, OR = 6.93; vs. LCIs: $p < .001$, OR = 7.78) odds of acceptance were significantly higher than the odds of the Spanish-dominant groups. While the IntHS had higher odds of accepting these items than the AdvHS (OR = 1.34), this difference in odds was not statistically significant, $p > .5$.

The fact that the HSs differ from the Spanish-dominant groups in their odds of acceptance with both Frequent and Infrequent verbs is noteworthy. This finding reveals

that Frequency effects alone cannot explain HSs' tendency to "overaccept" infelicitous Indicative mood items within Wrong scenario experimental items.

In the Subjunctive mood items, there was a significant effect of Group for both items with Frequent ($F(3,482) = 4.220$, $p < .01$) and Infrequent ($F(3,707) = 5.195$, $p < .01$) verbs. In the SubjunctiveFrequent condition, pairwise comparisons revealed that the SDCs' ($p < .01$, OR = 5.59) and the AdvHSs' ($p < .01$, OR = 4.47) odds of acceptance were significantly higher than the odds for the IntHSs. This finding is consistent with the explanation that the IntHSs do not strongly associate subjunctive mood with non-presuppositional contexts. In the SubjunctiveInfrequent condition, pairwise comparisons revealed that the SDCs' (vs. LCIs: $p < .05$, OR = 4.80; vs. IntHSs: $p < .01$, OR = 6.35) and AdvHSs' (vs. LCIs: $p < .01$, OR = 6.98; vs. IntHSs: $p < .001$, OR = 9.23) odds of acceptance were significantly higher than the odds for the other experimental groups. Interestingly, the LCIs show lower odds of acceptance than the SDCs with Subjunctive mood items, but only with Infrequent verbs. It is possible that the LCIs reject felicitous, SubjunctiveInfrequent items more because they are less familiar with the infrequent verb forms and therefore are not sure whether they are inflected with subjunctive or indicative morphology.

Figure 14 presents the results of the three-way interaction between Group, Mood and Frequency in GLMM #3, which focuses on the Wrong scenario items only.

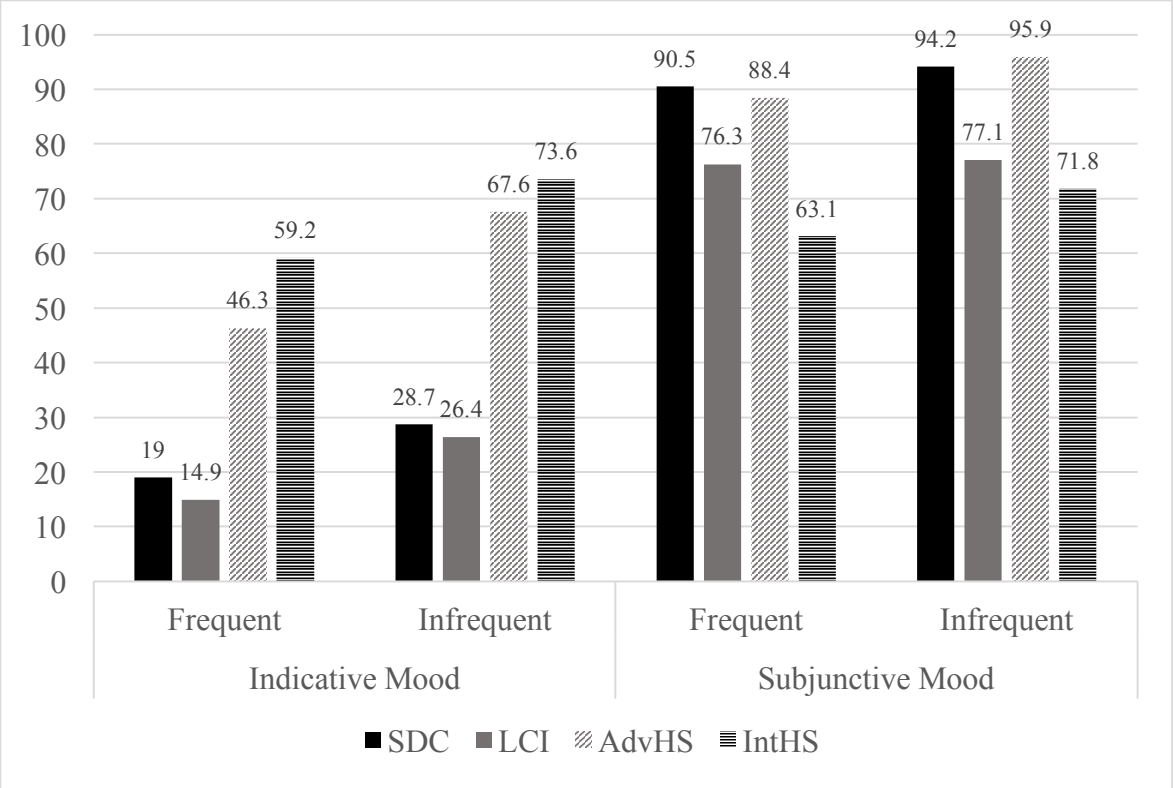

Figure 14: Probability of acceptance: Group x Frequency x Mood (Wrong)

Summary of three-way interaction

Analysis of the three-way interaction revealed the following trends. Unlike in the Found scenario items (GLMM #2), Frequency does not significantly impact any of the groups' odds of acceptance with either Indicative or Subjunctive mood items, respectively. Nonetheless, an examination of between-group differences reveals two important Frequency-related trends. First, in the Indicative mood items, the HSs' odds of acceptance differ significantly from the odds of the Spanish-dominant groups, regardless of Frequency. This indicates that Frequency does not explain why the HSs' are significantly more accepting of infelicitous Indicative forms in the Wrong scenario. In the Subjunctive mood items, we see an interesting Frequency effect in the performance of the LCIs, who are significantly less accepting of Subjunctive mood items with Infrequent verbs. This finding

suggests that the LCIs, despite a relatively later age of acquisition of English, may begin to lose sensitivity to mood with Infrequent verbs after prolonged English contact.

5.3.4. Summary of results from the CAT

The results of the CAT shed light on the three RQs in the present study, specifically as they relate to participants' *receptive* knowledge of subjunctive mood.

With respect to the first RQ, the results illustrate that HSs' receptive knowledge of polarity subjunctive mood is quite variable in nature. Compared to the Spanish-dominant controls, the HSs were more likely to accept infelicitous subjunctive items in the Found (+presupposition) scenario *and* more likely to accept infelictious indicative items in the Wrong (- presupposition) scenario. Clearly, the HS groups are less sensitive to mood distinctions in adjectival relative clauses.

The second RQ addresses the role of two between-group factors on participants' acceptability patterns. The results of the CAT suggest that Spanish-English bilinguals with later AofA Eng demonstrate greater sensitivity to mood distinctions than the HS groups, who began acquiring English at age 6 or earlier. This age effect is also apparent in the comparison between the SDCs, who acquired English at age 13 or later, and the LCIs, who began acquiring English between age 8 and 12. While the LCIs performed identically to the SDCs in the Wrong scenario items, they showed a much stronger tendency to accept Subjunctive mood items in the Found scenario, mirroring their pattern of "overproducing" subjunctive mood in the CEPT. Within the HS groups, proficiency also clearly modulated acceptability patterns. The AdvHSs, despite their lower sensitivity to mood distinctions, performed similarly to the SDCs and LCIs, accepting indicative items more in the Found scenario and subjunctive items more in the Wrong scenario. The IntHSs, on the other hand,

did not perform like the SDCs and LCIs, instead exhibiting an increased willingness to accept indicative items more regardless of scenario.

The third RQ addresses the effects of within-group factors on participants' knowledge of subjunctive mood. As with the results of the CEPT, the results from the CAT indicate that verb frequency modulates, at least to some extent, participants' knowledge of subjunctive. Generally speaking, participants were slightly more likely to accept items with infrequent verbs, perhaps due to an inability to recognize mood morphology on less frequent verbs. In the Found scenario items, the LCIs, AdvHSs and IntHSs were all more likely to accept infelicitous subjunctive mood items with infrequent verbs. A similar pattern was observed with the Wrong scenario items, where the AdvHSs and IntHSs were more likely to accept infelicitous indicative items with infrequent verbs. We will return to discuss the importance of these findings in Chapter 6.

5.4 Mood Preference Task (MPT)

In the MPT, participants listen to pairs of sentences that differ only in mood morphology. After listening to each sentence pair, they pick the sentence that sounds better.

The purpose of the MPT is to explore HSs' receptive knowledge of mood, thereby addressing RQ #1. By testing the role of both between-group (RQ #2) and within-group (RQ #3) factors on HSs' mood preferences, the MPT also addresses RQ's #2 and #3. Crucially, the MPT assesses receptive knowledge of both polarity and intensional mood selection, unlike the CAT which only tested polarity mood knowledge.

5.4.1 Descriptive statistics

In the subsections that follow, I will first present the variables of the MPT. Then, I will exhibit the raw data, which I will use to make two arguments. First, I will use the data from the Spanish-dominant controls to argue that the experimental design is valid. Second, I will briefly present the data from the fillers to argue that all experimental groups, including the HSs, are paying attention to the items in the MPT.

5.4.1.1. Variables in the MPT

Recall that in the MPT, participants select one sentence in each pair which "sounds better." Because there are clear target responses (e.g., participants are expected to prefer subjunctive mood after the complementizer, *para que*) in each pair, the dependent variable is Accuracy. Like the dependent variables in the CEPT and CAT, this dependent variable is binary: all responses are coded as either 1 ('accurate') or 0 ('inaccurate').

The independent variables in the MPT are Group (SDC, LCI, AdvHS or IntHS), ExpectedMood (Subjunctive or Indicative) and MoodSelectType (Intensional or Polarity). In the analyses that follow, I will demonstrate how each of these variables impacts participants' likelihood of responding accurately throughout the MPT.

5.4.1.2. Descriptive results

All 81 participants completed the MPT, which included 20 experimental items (4 conditions with 5 items each) and 10 fillers (2 conditions with 5 items each) for a total of 2430 responses (1620 experimental), all of which were coded for accuracy. All participants completed all items in the MPT, so no items were excluded from the analysis that follows.

Prior to presenting the statistical models for this task, I will first provide the raw data from the MPT. In revealing the descriptive data, I will achieve two goals. By highlighting the target performance of the SDCs, I will demonstrate the validity of the experimental task. By demonstrating the experimental groups' accuracy in the filler conditions, I will show that all experimental groups are paying attention during the MPT.

Recall that participants in the MPT listened to contexts and then heard two follow-up sentences, one with indicative and another with subjunctive morphology. Participants' job was to select the follow-up sentence that sounded better. All responses were coded '1' (accurate) or '0' (inaccurate), based on the expected target mood-form for each item.

As shown in Table 39, the SDCs perform as expected with an average accuracy of at least 92% in each of the four experimental conditions. In the IntensionalSubjunctive condition, the SDCs were 100% accurate. Because of the SDCs' invariant performance in this condition, we will have to slightly modify the data in order to run GLMMs in Section 5.4.3.1. Not surprisingly, the LCIs also perform quite accurately in all conditions while the AdvHSs and IntHSs perform more variably throughout the MPT.

Table 39.
Response accuracy by Group, MoodSelectType, ExpectedMood

Group	Mood Select Type			
	Intensional		Polarity	
	Indicative *Porque*	Subjunctive *Para que*	Indicative Presupposition	Subjunctive -Presupposition
SDCs	Mean: 95.0 SD: 21.9	Mean: 100.0 SD: 0	Mean: 95.0 SD: 21.9	Mean: 92.0 SD: 27.3
LCIs	Mean: 93.7 SD: 24.4	Mean: 97.9 SD: 14.4	Mean: 90.5 SD: 21.4	Mean: 91.6 SD: 27.9
AdvHSs	Mean: 90.9 SD: 28.9	Mean: 91.8 SD: 27.5	Mean: 82.7 SD: 38.0	Mean: 66.4 SD: 47.5
IntHSs	Mean: 83.0 SD: 37.8	Mean: 78.0 SD: 41.6	Mean: 80.0 SD: 40.2	Mean: 46.0 SD: 50.1

All participants completed the MPT at the very end of the experiment, at which point fatigue may have affected their ability to pay attention to the task. To show that participants in all groups were paying (at least some) attention to the experimental items, we now examine the filler items for the MPT, which evaluated participants' knowledge of both argument structure (*Busco* vs. **Busco por*) and verbal agreement (*Nos gustaron vs.* **Nos gustó* with plural themes) with the reverse psychological verb *gustar*.

Table 40.
Response accuracy by Group, Condition (Fillers)

	Buscar Fillers	*Gustar* Fillers
SDCs	Mean: 91.6 SD: 16.7	Mean: 86.3 SD: 25.0
LCIs	Mean: 86.3 SD: 26.7	Mean: 58.9 SD: 30.2
AdvHSs	Mean: 89.1 SD: 25.2	Mean: 63.4 SD: 35.8
IntHSs	Mean: 76.0 SD: 30.1	Mean: 59.0 SD: 34.6

As shown in Table 40, all groups, including the HSs, were highly accurate with the *Buscar* fillers, indicating that they were paying attention to the experimental task. With the *Gustar* fillers, on the other hand, only the SDCs (86.3%) performed with high accuracy. The LCIs and the HS groups may perform less accurately with these fillers due to changing agreement patterns in US Spanish (Pascual y Cabo, 2013, 2016), rather than a lack of attention to the experimental stimuli.

5.4.2. Statistical models for the MPT

Because the dependent variable is binary (e.g., accurate or inaccurate), we must use GLMMs, rather than Mixed ANOVA's, to analyze the data. As we have seen in Sections 5.2 and 5.3, GLMMs allow us to estimate how different independent variables influence

participants' probability of responding in a certain way, e.g., accurately or inaccurately. Before demonstrating the results of the GLMM for the MPT, however, we must first briefly discuss how the data was transformed and prepared for analysis with GLMMs.

5.4.2.1 Transforming the data

As discussed in Section 5.2., GLMMs cannot operate when data within a single cell (e.g., Group A in Condition 1) is invariant, that is to say, exhibiting a certain behavior either 0% or 100% of the time. In Section 5.4.1.2, we saw that the SDCs were 100% accurate in their responses to the Intensional, ExpectedSubjunctive condition. Consequently, the researcher selected two of the SDC responses in this condition (one from Version A, another from Version B) and changed them from 1 ('accurate') to 0 ('inaccurate'). This introduced necessary variance into the data set and allowed for the possibility of using GLMMs throughout the analysis.

5.4.3. GLMMs

In the section that follows, I will use GLMMs to analyze the results of the MPT. As in Sections 5.2. and 5.3, I will present the GLMMs in the following manner. The independent variables will be presented, in all capital letters, as part of the section heading. Within the analysis itself, I will present main effects, as well as two- and three-way interaction effects. To do so, I will use tables to present participants' log-odds, standard errors, and odds (of responding accurately), as well as their predicted probabilities (with 95% confidence intervals.) To conclude the analysis, I will use a bar graph to provide a visual representation of the three-way interactions.

5.4.3.1. Group, MoodSelectType and ExpectedMood

The data were analyzed with a GLMM. The independent variables were Group (AdvHS, IntHS, SDC, LCI), MoodSelectType (Intensional, Polarity) and ExpectedMood (Subjunctive, Indicative) and the dependent variable was Accuracy (Accurate or Inaccurate). The model also included random subject and item intercepts, which allow the researcher to control for random variance among subjects and lexical items.

5.4.3.1.1. Main effects

This GLMM (GLMM #1) revealed significant main effects of Group ($F(3,95) = 12.763$, $p < .001$) and MoodSelectType ($F(1,18) = 7.267$, $p < .05$) but no significant effect of ExpectedMood ($F(1,18) = .393$, $p > .5$). Pairwise comparisons for Group, highlighted in Table 41, revealed that overall, the SDCs (vs. AdvHS: $p < .01$, OR = 3.95; vs. IntHS: $p < .001$, OR = 8.50) and LCIs (vs. AdvHS: $p < .01$, OR = 3.04; vs. IntHS: $p < .001$, OR = 6.54) were significantly more likely to respond accurately than the HS groups. Within the heritage groups, the AdvHSs' odds of responding accurately were significantly higher than the odds for the IntHSs, $p < .05$, OR = 2.15. There were no significant differences between the SDCs' and the LCIs' predicted probability of responding accurately.

Table 41.
Probability of accuracy by Group

Group	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
SDC	3.331	.371	27.966	96.5%	93.1%	98.3%
LCI	3.069	.362	21.520	95.6%	91.3%	97.8%
AdvHS	1.957	.289	7.078	87.6%	79.9%	92.7%
IntHS	1.191	.282	3.290	76.7%	65.1%	85.3%

Pairwise comparisons for MoodSelectType, shown in Table 42, revealed that in general, participants' odds of accurately responding to a given experimental item were significantly higher for Intensional items than for Polarity items, $p < .05$, OR = 2.82.

Table 42.
Probability of accuracy by MoodSelectType

Mood Select Type	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
Intensional	2.906	.304	18.284	94.8%	90.7%	96.6%
Polarity	1.869	.274	6.482	86.6%	78.5%	92.0%

Though there was no significant main effect of ExpectedMood ($p > .5$), as shown in Table 43, participants showed slightly higher odds (OR = 1.27) of responding accurately when Indicative, as opposed to Subjunctive, was the expected mood.

Table 43.
Probability of accuracy by ExpectedMood

Expected Mood	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
Subjunctive	2.267	.295	9.650	90.6%	84.0%	94.7%
Indicative	2.508	.284	12.280	92.5%	87.2%	95.7%

Summary of main effects

GLMM #1 revealed two statistically significant main effects. First, the Spanish-dominant groups were significantly more likely to respond accurately than the HS groups, though each of the HS groups performed above 75% accuracy. Second, participants in general were more likely to respond accurately with Intensional mood selection.

5.4.3.1.2. Two-way interactions

GLMM #1 revealed a statistically significant two-way interaction between Group and ExpectedMood, $F(3,1604) = 3.916$, $p < .01$, but no significant interaction between

either MoodSelectType and ExpectedMood (F (1,18) = 2.384, $p > .1$) or Group and MoodSelectType (F(3,1604) = 0.505, $p > .6$).

We start by considering the Group by ExpectedMood interaction, shown in Table 44, which addresses two questions. First, do participants in each group show higher odds of responding accurately when the ExpectedMood is Indicative? Results of an additional GLMM reveal that only the IntHSs ($p < .05$, OR = 3.12) were more likely to be accurate in conditions where Indicative is the ExpectedMood. The AdvHSs' ($p > .2$, OR = 1.68), SDCs' ($p > .8$ OR = 0.88) and LCIs' ($p > .3$, OR = 0.57) odds of responding accurately were not significantly affected by ExpectedMood.

Table 44.
Probability of accuracy: Group x ExpectedMood

Group	Exp.Mood	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
SDC	Subjunctive	3.395	.510	29.815	96.8%	91.6%	98.8%
	Indicative	3.267	.454	26.233	96.3%	91.4%	98.5%
LCI	Subjunctive	3.350	.512	28.503	96.6%	91.2%	98.7%
	Indicative	2.788	.420	16.248	94.2%	87.6%	97.4%
AdvHS	Subjunctive	1.699	.359	5.254	84.0%	72.6%	91.8%
	Indicative	2.215	.369	9.161	90.1%	81.4%	95.0%
IntHS	Subjunctive	.622	.340	1.863	65.1%	48.3%	78.8%
	Indicative	1.760	.359	5.812	85.3%	73.8%	92.3%

The between-groups perspective of this interaction answers a second question, namely whether there are significant differences between the groups in each of the levels of ExpectedMood. An additional GLMM revealed a significant effect of Group both when Indicative is expected (F(3,178) = 3.957, $p < .01$) and when Subjunctive is expected (F(3,215) = 14.848, $p < .001$). When Indicative was expected, the SDCs' odds of responding accurately were significantly higher than the odds for the IntHSs ($p < .01$, OR = 4.51) and the AdvHSs ($p < .05$, OR = 2.86) but not the LCIs ($p > .3$, OR = 1.61). The LCIs' odds of responding accurately were significantly higher than the odds for the IntHSs

only, $p < .05$, OR = 2.80. It is somewhat surprising to see that the HSs diverge from the Spanish-dominant groups in the ExpectedIndicative condition, where we would not expect the HSs to exhibit variability. This result may be driven by HSs' lexical instability, which affects both indicative, as well as subjunctive, inflection.

When Subjunctive was expected, the SDCs' and LCIs' odds of responding accurately were significantly higher than the odds for the AdvHSs (SDCs: $p < .01$, OR = 5.45; LCIs: $p < .01$, OR = 5.21) and the IntHSs (SDCs: $p < .001$, OR = 16.01; LCIs: $p < .001$, OR = 15.30). Unlike in the ExpectedIndicative condition, where there were no significant differences between the HS groups, the AdvHSs' were more likely to respond accurately than the IntHSs when Subjunctive was expected, $p < .01$, OR = 2.94.

Recall that the Group x MoodSelectType interaction, shown below in Table 45, was not statistically significant ($p > .1$), largely because all participant groups showed slightly higher odds of accuracy with Intensional mood selection. Nonetheless, there is still value in examining (a) within-group and (b) between-group pairwise comparisons for this two-way interaction. We start by examining the within-group pairwise comparisons, which tell us whether each participant group has higher odds of responding accurately with Intensional versus Polarity mood selection.

An additional GLMM showed that for the SDCs ($F(1,121) = 1.628$, $p > .2$, OR = 2.20) and the LCIs ($F(1,102) = 2.954$, $p > .08$, OR = 2.77), the odds of responding accurately were not significantly higher for Intensional, as opposed to Polarity mood selection items. While these differences were not statistically significant, it is noteworthy that the odds of responding accurately for both of the Spanish-dominant groups are over two times greater for Intensional items than for Polarity items. For the AdvHS ($F(1,32) =$

9.682, $p < .01$, OR = 3.97) and IntHS (F(1,24) = 5.398, $p < .05$, OR = 2.61), however, the odds of responding accurately are significantly higher with Intensional as opposed to Polarity mood selection. This result complements the findings of the CEPT, in which both AdvHSs and IntHSs were more likely to *produce* Intensional subjunctive.

Table 45.
Probability of accuracy: Group x MoodSelectType

Group	Exp.Mood	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
SDC	Intensional	3.726	.531	41.513	97.6%	93.6%	99.2%
	Polarity	2.937	.429	18.859	95.0%	88.9%	97.8%
LCI	Intensional	3.579	.525	35.838	97.3%	92.7%	99.0%
	Polarity	2.559	.404	12.923	92.8%	85.2%	96.7%
AdvHS	Intensional	2.647	.389	14.112	93.4%	86.6%	96.8%
	Polarity	1.268	.338	3.554	78.0%	64.2%	87.6%
IntHS	Intensional	1.671	.357	5.317	84.2%	72.1%	91.6%
	Polarity	.711	.343	2.036	67.1%	50.4%	80.3%

The two-way interaction between MoodSelectType and ExpectedMood, shown in Table 46, was also not statistically significant ($p > .6$). However, further analysis of this interaction still offers considerable insight into the effect of MoodSelectType on participants' odds of accurately responding in the MPT. Results of an additional GLMM revealed a significant effect of MoodSelectType when Subjunctive was expected (F(1,20) = 8.631, $p < .01$, OR = 5.11) but not when Indicative was expected (F(1,17) = .692, $p > .4$, OR = 1.56). Subsequent pairwise comparisons showed that when Subjunctive was expected, participants' odds of responding accurately were significantly higher (by a factor of 5) with Intensional, as opposed to Polarity, mood selection, $p < .01$.

This result clearly indicates that participants' odds of responding accurately are higher with Intensional mood selection than with Polarity mood selection, though, crucially, this effect appears only when subjunctive is the expected mood. This is exactly

the effect we would expect to see if Indicative is the default mood form and Intensional subjunctive forms are "easier" for participants to acquire and/or access from memory.

Table 46.
Probability of accuracy: ExpectedMood x MoodSelectType

Exp. Mood	Select Type	Logodds	S. Error	Odds	Probability	CI Lower	CI Upper
Indicative	Intensional	2.729	.397	15.318	93.9%	87.0%	97.2%
	Polarity	2.286	.382	9.386	90.8%	81.5%	95.6%
Subjunctive	Intensional	3.082	.438	21.802	95.6%	90.0%	98.2%
	Polarity	1.451	.368	4.267	81.0%	66.1%	90.3%

Summary of two-way interactions

GLMM #1 revealed a significant interaction between Group and ExpectedMood. Out of the experimental groups, only the IntHSs are more likely to respond accurately when Indicative mood is expected. The between-groups perspective of this interaction revealed that the HS groups' odds of responding accurately are significantly different from the odds of the Spanish-dominant groups, even when Indicative is the ExpectedMood, suggesting underlying lexical instability in the HS groups. An analysis of the Group by MoodSelectType interaction revealed that the HS groups, but not the Spanish-dominant groups, were more likely to respond accurately with Intensional mood selection. Complementing this finding is the analysis of the interaction between ExpectedMood and MoodSelectType, which revealed a facilitation effect for Intensional mood selection, but only when ExpectedMood is subjunctive. To complete our understanding of the interactions between these three variables, we now consider the three-way interaction.

5.4.3.1.3. Three-way interaction

The GLMM revealed that the Group x ExpectedMood x MoodSelectType interaction (Table 47) was not statistically significant, $F(3,1604) = .096$, $p > .9$.

Nonetheless, the three-way interaction still allows us to make a number of pairwise comparisons which will prove invaluable to the discussion and analysis of the MPT.

Table 47.
Probability of accuracy: Group x ExpectedMood x MoodSelectType

Group	Exp. Mood	Mood Selection	Logodds	SE	Odds	Probability	CI-Lower	CI-Upper
SDC	Subjunctive	Intensional	4.143	.818	62.992	98.4%	92.6%	99.7%
		Polarity	2.648	.536	14.126	93.4%	82.9%	97.6%
	Indicative	Intensional	3.310	.612	27.385	96.5%	89.0%	98.9%
		Polarity	3.225	.606	25.154	96.2%	88.3%	98.8%
LCI	Subjunctive	Intensional	4.098	.819	60.220	98.4%	92.3%	99.7%
		Polarity	2.602	.539	13.491	93.1%	82.1%	97.5%
	Indicative	Intensional	3.060	.586	21.328	95.5%	86.9%	98.6%
		Polarity	2.516	.528	12.379	92.5%	81.2%	97.3%
AdvHS	Subjunctive	Intensional	2.645	.518	14.083	93.4%	83.3%	97.5%
		Polarity	.754	.425	2.125	68.0%	47.0%	83.6%
	Indicative	Intensional	2.649	.514	14.140	93.4%	83.5%	97.5%
		Polarity	1.782	.457	5.942	85.6%	70.1%	93.8%
IntHS	Subjunctive	Intensional	1.443	.454	4.233	80.9%	62.7%	91.4%
		Polarity	-.200	.430	.819	45.0%	25.3%	66.4%
	Indicative	Intensional	1.899	.476	6.679	87.0%	71.8%	94.6%
		Polarity	1.621	.462	5.058	83.5%	66.5%	92.8%

We begin our analysis of this interaction by examining between-group differences in the odds of responding accurately in each of the four experimental conditions: IntensionalSubjunctive, IntensionalIndicative, PolaritySubjunctive and PolarityIndicative.

In the IntensionalSubjunctive condition, a GLMM revealed a significant effect of Group ($F(3,870) = 6.501$, $p < .001$) on participants' odds of responding accurately. Pairwise comparisons showed that the SDCs ($p < .01$, OR = 14.88), LCIs ($p < .01$, OR = 14.22), and AdvHS ($p < .05$, OR = 3.33) all had significantly higher odds of responding accurately than the IntHS. Both the SDCs ($p > .07$, OR = 4.47) and the LCIs ($p > .08$, OR = 4.28) had marginally higher odds of responding accurately than the AdvHS, though this difference was not statistically significant.

Not surprisingly, there was an even stronger significant effect of Group in the PolaritySubjunctive condition, $F(3,244) = 15.313$, $p < .001$. Pairwise comparisons revealed that both the SDCs (vs. AdvHS: $p < .001$, OR = 6.65; vs. IntHS: $p < .001$, OR = 17.25) and the LCIs (vs. AdvHS: $p < .001$, OR = 6.35; vs. IntHS: $p < .001$, OR = 16.48) had significantly higher odds of responding accurately than the HS groups. Within the HS groups, there was a substantial proficiency effect, as the AdvHS' odds of responding accurately were significantly higher than the odds for the IntHS, $p < .05$, OR = 2.60.

Thus far, we have seen considerable differences between groups in the two conditions of the MPT in which Subjunctive mood is expected. If indicative mood morphology is default, either in general or for the HS groups specifically, we might expect to find much smaller group differences in the IntensionalIndicative and PolarityIndicative conditions, respectively. To consider this possibility, we now turn to between-group comparisons of the two IndicativeExpected conditions.

In the IntensionalIndicative condition, a GLMM revealed a marginally significant effect of Group, $F(3,545) = 2.292$, $p = .07$. However, pairwise comparisons revealed only one statistically significant difference between groups: the SDCs were significantly more likely to respond accurately than the IntHS, $p < .05$, OR = 4.10, a finding which reflects the IntHS' tendency to occasionally acccept subjunctive with *porque*. (Note that this tendency mirrors the IntHSs' occasional tendency to produce subjunctive with *porque* in the CEPT.) In the PolarityIndicative condition, another GLMM revealed a significant effect of Group, $F(3,365) = 2.966$, $p < .05$. Pairwise comparisons indicated that the odds of the SDCs responding accurately in this condition were higher than the odds for the AdvHS ($p < .05$, OR = 4.23) and the IntHS ($p < .01$, OR = 4.97).

It is critical to supplement *between-group* comparisons, which compare HSs to Spanish-dominant groups, with *within-group* comparisons, which compare HSs to *themselves*. In this vein, let's now consider (a) whether participants in each group have higher odds of responding accurately with Intensional or Polarity mood selection and (b) whether this potential difference in odds is conditioned by ExpectedMood.

When Indicative is expected, there appears to be minimal impact of MoodSelectType on participants' odds of responding accurately. For the SDCs ($F(1,88) = .011$, $p > .9$, OR = 1.09), LCIs ($F(1,59) = .552$, $p > .4$, OR = 1.72), AdvHS ($F(1,33) = 1.862$, $p > .1$, OR = 2.38) and IntHS ($F(1,27) = .212$, $p > .6$, OR = 1.32), the odds of responding accurately with ExpectedIndicative items do not differ based on MoodSelectType. When Subjunctive is expected, however, MoodSelectType appears to make a stronger impact on all of the experimental groups. Both the SDCs ($F(1,159) = 2.559$, $p > .1$, OR = 4.46) and LCIs ($F(1,159) = 2.565$, $p > .1$, OR = 4.46) show a marked, though non-statistically significant, tendency to respond more accurately with Intensionally selected items. For the AdvHS ($F(1,30) = 9.383$, $p < .01$, OR = 6.63) and the IntHS ($F(1,21) = 8.463$, $p < .01$, OR = 5.17), the odds of responding accurately in the ExpectedSubjunctive conditions are significantly higher with Intensionally selected items.

Just as MoodSelectType seems to impact accuracy more when Subjunctive mood is expected, so ExpectedMood impacts accuracy more on items with Polarity Mood selection. With IntensionalSelection, it appears that there is little apparent impact of ExpectedMood on participants' odds of responding accurately: SDCs ($F(1,187) = .724$, $p > .3$, OR = 2.30), LCIs ($F(1,174) = 1.165$, $p > .2$, OR = 2.82), AdvHS ($F(1,44) = .000$, $p > .9$, OR = 1.00), and IntHS ($F(1,27) = .581$, $p > .4$, OR = 0.63). With PolaritySelection,

however, there is a significant impact of ExpectedMood on odds of responding accurately, though only for the HS groups. For the SDCs (F(1,68) = .581, $p > .4$, OR = 1.78) and the LCIs (F(1,49) = .015, $p > .5$, OR = .92), ExpectedMood does not significantly impact accuracy. For the AdvHS (F(1,22) = 3.278, $p < .09$, OR = 2.80) and the IntHS (F(1,22) = 10.183, $p < .01$, OR = 6.18), however, ExpectedMood plays a significant role. It is very noteworthy that the IntHS' odds of responding accurately increase sixfold in Polarity Mood conditions when Expected Mood is indicative, rather than subjunctive.

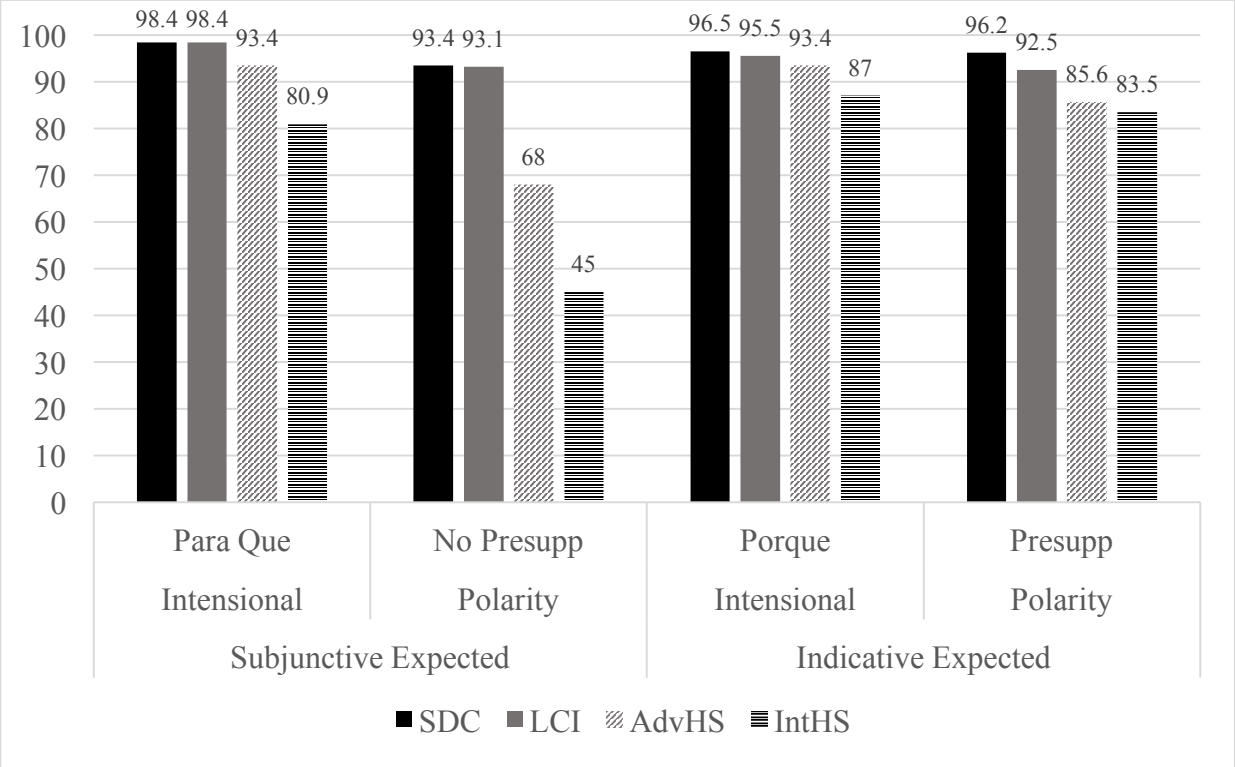

Figure 15: Probability of accuracy: Group x ExpectedMood x MoodSelectType

Figure 15 provides a graphical representation of the three-way interaction between Group, ExpectedMood and MoodSelectType.

Summary of three-way interaction

Analysis of the three-way interaction revealed the following trends. From a between-group perspective, GLMM #1 revealed significant effects of Group on participants' odds of responding accurately in each of the four MPT conditions. The IntHSs differed from the Spanish-dominant groups in all four conditions while the AdvHSs differed from the Spanish-dominant groups only in the two conditions with Polarity mood selection. The model also revealed a substantial proficiency effect in each of the SubjunctiveExpected conditions, whereby the AdvHSs' odds of responding accurately were significantly higher than the odds for the IntHSs.

From a within-group perspective, the GLMM revealed a significant effect of MoodSelectType but only for the HSs and only when ExpectedMood is subjunctive. There was also a significant effect of ExpectedMood for the HSs, but only with Polarity selection.

5.4.4. Summary of results from the MPT

The results of the MPT directly inform the RQs of the study. The first RQ addresses the nature of HSs' subjunctive knowledge. Results from the MPT, like the results of the CEPT and CAT, indicate that HSs' knowledge of subjunctive mood is quite variable in nature. Overall, HSs' mood preferences in the MPT are significantly less accurate than the preferences of the Spanish-dominant controls and LCIs. Despite this variability, however, the HSs perform more accurately in the MPT, which gauges their *receptive* knowledge of subjunctive, than in the CEPT, which assessed their *productive* subjunctive knowledge. In the IntensionalSubjunctive condition, for example, both the AdvHSs (93.4% accuracy in MPT, cf 86.6% accuracy in CEPT) and the IntHSs (80.9% accuracy in MPT, cf 47.4% accuracy in CEPT) perform more accurately in the preference task than in the

Contextualized Elicited Production Task. I will return to explore the significance of this finding in Chapter 6.

The second RQ addresses the role of between-group variables, specifically, AofA Eng, and Spanish proficiency, on participants' subjunctive knowledge. The results of the MPT, like the results of the CEPT and CAT, reinforce the critical role of AofA Eng. Again, the HS groups performed less accurately than both the SDCs (age of acquisition of English = 12+) and the LCIs (age of acquisition of English between 8-12). Surprisingly, the HS groups performed less accurately regardless of the expected mood. I will return to explore this counterintuitive finding in Chapter 6. AofA Eng did not, however, lead to differences in accuracy between the SDCs and the LCIs, both of whom reached 90% accuracy or above in all four experimental conditions, including the PolarityIndicative condition, where the LCIs had previously exhibited a stark tendency to both "overproduce" and "overaccept" subjunctive mood forms. In the case of the HSs, the between-group factor of Spanish proficiency played an important role in participants' accuracy. The AdvHSs were more accurate overall than the IntHSs, due to their significant accuracy advantage in the SubjunctiveExpected conditions.

Finally, the third RQ addressed the role of within-group factors on participants' knowledge of subjunctive mood. Results of the MPT indicate that both the AdvHSs and IntHSs demonstrate a clear effect of MoodSelectType. Specifically, when subjunctive mood is expected, both the HS groups respond more accurately with Intensional mood selection than with Polarity mood selection. This pattern closely mirrors the HS groups' performance on the CEPT, where they were more likely to produce Intensional than Polarity subjunctive mood forms.

5.5. Summary of Chapter 5

In this chapter, I have presented the results from the Contextualized Elicited Production Task (CEPT), the Contextualized Acceptability Task (CAT), and the Mood Preference Task (MPT). In presenting the results of these tasks, I have addressed the RQs of the present study, shedding light on the nature of HSs' subjunctive knowledge (RQ#1) as well as various between-group (RQ #2) and within-group (RQ #3) variables that affect that knowledge.

In addition, I have shown that HSs, despite diverging from the SDCs, often exhibit qualitatively similar grammatical knowledge of subjunctive mood morphology in Spanish. In Chapter 6, I will review these findings, focusing primarily on their implications for the three accounts of HS divergence/variability presented in Chapter 2.

CHAPTER 6:
DISCUSSION

In Chapter 5, I presented the results from the Contextualized Elicited Production Task (CEPT), the Contextualized Acceptability Task (CAT), and the Mood Preference Task (MPT). In this chapter, I will explore the relevance of these findings both for the RQs presented in Chapter 4, as well as for HL acquisition research more generally.

The remainder of Chapter 6 will be organized as follows. In Section 6.1, I will provide a broad overview of the results, highlighting the presence of both HS divergence and HS variability, respectively. In Sections 6.2 and 6.3., I will build on this broad analysis by digging into factors which affect whether, and to what extent, HS divergence and variability are observed. In Section 6.2, I will use group, as well as individual results, to summarize how different *between-group* factors shape HSs' divergence with subjunctive mood. In Section 6.3., I will utilize group and individual results to address the role of different *within-group* factors modulating HSs' subjunctive mood variability. Throughout both of these sections, I will draw comparisons between the three approaches to HS divergence/variability presented in Chapter 2, suggesting that the Activation/Lexicalist approach most adequately accounts for the patterns observed. Finally, in Section 6.4, I will conclude the chapter by presenting a unified account of the different ways in which HSs come to exhibit divergence and variability with subjunctive mood in Spanish.

6.1. HS differences: divergence and variability

The first RQ, copied below from Chapter 4, was descriptive, rather than explanatory, addressing the observable characteristics of HSs' subjunctive mood usage.

1. What is the nature of HSs' knowledge of lexically (intensional) and contextually (polarity) selected mood morphology in Spanish?

In Chapter 2, I introduced, defined, and defended two different terminological tools which we can use to label differences in the nature of HSs' knowledge of a given property. HS divergence, which I will address in Section 6.1.1., is when HSs differ, either qualitative- or quantitatively, from dominant-speakers of the HL, e.g., producing a certain target grammatical form less often. HS variability, which I will discuss in Section 6.1.2, is when HSs differ from *themselves*, e.g., alternating between target- and non-target variants of a single grammatical form. As we saw in Chapter 5, HSs exhibit both divergence and variability in their knowledge of subjunctive mood.

6.1.1. HS divergence

The results of this study reveal clear evidence of HS divergence in both the AdvHS and IntHS groups, respectively, across both productive and receptive experimental tasks.

In the CEPT, the HSs diverged quantitatively from the Spanish-dominant controls (SDCs), producing significantly less subjunctive mood with both intensional and polarity mood selection. If we assume that the SDCs' categorical production of subjunctive in these conditions approximates the subjunctive production of the HSs' parents, then it is clear that the HSs' subjunctive production diverges strongly from the input they received.

In the CAT, which tested HSs' sensitivity to polarity mood alternations in adjectival relative clauses, the HSs once again diverged from the SDCs, although the type and extent of this divergence was modulated by proficiency. Like the SDCs, the AdvHSs showed higher acceptance of indicative mood forms in presuppositional contexts and subjunctive

mood forms in non-presuppositional contexts, thus mirroring the qualitative distinctions made by the SDCs. Nonetheless, the AdvHSs diverged quantitatively from the SDCs, showing a much higher tendency to accept both (a) infelicitous subjunctive mood forms in presuppositional contexts and (b) infelicitous indicative mood forms in non-presuppositional contexts. The IntHSs, on the other hand, differed both quantitatively and qualitatively from the SDCs. In addition to showing much higher acceptance of infelicitous subjunctive and indicative mood items, the IntHSs also demonstrated a lack of sensitivity to mood in non-presuppositional relative clauses. Unlike the SDCs and AdvHSs, both of whom showed higher acceptance of subjunctive mood in these clauses, the IntHSs accepted both subjunctive (target) and indicative (non-target) forms equally, thereby diverging qualitatively, as well as quantititavely, from the performance of the SDCs.

In the MPT, which tested participants' mood preferences with both intensional and polarity selection, the HSs again diverged from the SDCs, though this divergence was modulated by expected mood, mood selection type, and proficiency. With intensional indicative items, only the IntHSs exhibited divergence from the SDCs, showing a higher tendency to prefer subjunctive mood with *porque*. With polarity indicative items, both the IntHSs and AdvHSs diverged from the SDCs, this time demonstrating a stronger tendency to prefer subjunctive in presuppositional contexts. Similar patterns of divergence emerged with expected subjunctive items of the MPT. With intensional subjunctive items, only the IntHSs exhibited divergence from the SDCs, due to their increased preference for indicative after *para que*. With polarity subjunctive items, however, both the IntHSs and AdvHSs diverged from the SDCs, as both groups showed a markedly lower tendency to prefer subjunctive mood in non-presuppositional contexts, as compared to the SDCs.

In summary, the HSs diverged from the SDCs in production, acceptance, and preference of subjunctive mood forms in Spanish. To answer the first part of RQ #1, therefore, we can state that the nature of HSs' productive and receptive knowledge of subjunctive mood is *divergent*. Across three different tasks, the HS groups' productive and receptive knowledge of mood differs from that of the SDCs, whose performance approximates the input that HSs likely received at home from their parents.

Before continuing on to Section 6.1.2, it is important to briefly discuss a surprising, and as yet unidentified, trend in the polarity mood data, which may have implications for our descriptive knowledge of polarity mood in adjectival relative clauses. In the CEPT, the SDCs (10.7%) and the LCIs (42.2%) both produced surprisingly high proportions of subjunctive mood in presuppositional adjectival relative clauses. In the CAT, the SDCs (29.7%) and the LCIs (52.6%) showed a pronounced, and once again, unexpected, tendency to also *accept* infelicitous subjunctive mood morphology in presuppositional contexts. Considered together, these seemingly anomalous productive and receptive tendencies appear to indicate that mood in Spanish adjectival relative clauses may be more complex than what has been previously described in the syntactic literature (e.g., Pérez-Leroux, 1998; Borgonovo et al, 2015). That is to say, there may be certain, as yet unidentified, *types* of adjectival relative clauses that include presupposition and yet nonetheless favor (or allow for) the production of subjunctive mood.

6.1.2. HS variability

In Section 6.1.1., I argued that HSs' knowledge of subjunctive mood is *divergent* because it differs, in measurable ways, from that of the SDCs. In order to fully characterize

the nature of HSs' subjunctive knowledge, however, it is equally important to consider the qualitative and quantitative nature of the HS grammars themselves. In this section, rather than comparing the HSs to the SDCs, I provide an overview of the variation demonstrated by HSs in their own subjunctive systems. I will argue, following Chapter 2, that HSs exhibit *variability* in their knowledge of subjunctive. In production, acceptance and preference, the HSs appear to demonstrate a less categorical mood system, occasionally producing, accepting and preferring indicative mood forms in contexts where SDCs produce, accept and prefer subjunctive mood forms categorically.

Though the HSs diverged from the SDCs on the CEPT, it is not the case that these groups diverged by not producing subjunctive mood forms *at all*. Quite on the contrary, the HS groups' performance was marked by a strong tendency to alternate, to differing degrees, between producing target subjunctive and non-target indicative forms in expected subjunctive conditions. The AdvHSs, on one hand, exhibited *minimal* mood variability in expected subjunctive conditions, producing mostly target subjunctive forms (86.6% with intensional selection; 74.3% with polarity selection) which alternated with occasional, non-target indicative inflections. The IntHSs also exhibited mood variability, albeit to a much greater extent than the AdvHSs. With intensional mood selection (47.4%), the IntHSs demonstrated *intensive* variability, producing target subjunctive and non-target indicative mood forms with nearly equal frequency. With polarity subjunctive (24.7%), however, the IntHSs exhibited only *moderate* variability, producing predominantly non-target indicative forms which alternated with occasional subjunctive forms.

As pointed out in Chapter 5, the fact that HSs exhibit variability in their production of subjunctive does not imply an *absence* of subjunctive knowledge. The AdvHSs, as it

turns out, were much more likely to produce subjunctive in expected subjunctive conditions (e.g., with *para que* and *without* presupposition) than in expected indicative conditions (e.g., with *porque* and *with* presupposition), and by a very substantial margin. Given that the AdvHSs exhibited only *minimal* variability in the CEPT, it is perhaps not surprising that they make such distinctions. That said, the IntHSs, who exhibited *moderate* variability with intensional subjunctive and *intensive* variability with polarity subjunctive *also* produced more subjunctive mood when subjunctive was expected, making surprisingly strong across-condition distinctions (all OR's > 10.00). In light of these findings, it is safe to say that HSs' *productive* variability, though different in nature from the categorical production of the SDCs, implies neither a random nor unsystematic mood knowledge.

In the CAT, which tested participants' receptive knowledge of polarity mood distinctions, HSs again showed evidence of grammatical variability. In Section 6.1.1., I noted that HSs were more likely than the SDCs to accept both (a) infelicitous subjunctive items and (b) infelicitous indicative items. When we consider this observation from the perspective of the HSs, it is apparent that HS grammars are clearly more variable, given that HSs accept, to a considerable extent, both indicative *and* subjunctive mood items in both presuppositional *and* non-presuppositional contexts.

For the IntHSs, who accept both indicative and subjunctive mood forms in non-presuppositional contexts, it appears that polarity subjunctive and indicative mood forms occur in free variation, which is the ultimate example of HS variability. It is not the case, however, that mood variability in the CAT is limited to the IntHSs. Even for the AdvHSs, who pattern with the SDCs in showing higher acceptance of (a) indicative with presupposition and (b) subjunctive without presupposition, it is clear that infelicitous and

felicitous indicative and subjunctive mood forms compete, to some extent, in their mental grammars. (If this were not the case, then why would the AdvHSs accept infelicitous subjunctive (56.5%) and infelicitous indicative (57.4%) items with such high frequency?). Based on the results of the CAT, it is apparent that HSs can and do exhibit variable comprehension, as well as production, of subjunctive mood, even when they appear to make the same across-condition distinctions as the SDCs.

The MPT offers a third, and perhaps most compelling, opportunity to observe variability in HSs' subjunctive mood knowledge. Recall that participants in the MPT listened to contexts, followed by a minimal pair of sentences which differed only in the mood of the verb in the subordinate clause (subjunctive or indicative). After listening to each pair of sentences, participants selected the sentences that sounded best to them. Variability in the MPT, therefore, is evident when HSs prefer an indicative mood form when a subjunctive mood form is expected, or vice versa. While the AdvHSs (93.4%) *and* the IntHSs (80.9%) prefer subjunctive mood forms with *para que*, only the AdvHSs (68.0%) prefer subjunctive in non-presuppositional contexts. (The IntHSs, on the other hand, prefer subjunctive just 45.0% of the time in these contexts.)

The results of the MPT provide particularly potent evidence of HSs' mood variability. Take, for instance, the case of the IntHSs' performance with *para que*. The fact that the IntHSs were 80.9% accurate in this condition means that approximately 20% of the time, they selected non-target indicative mood items with *para que*, even though they heard the target subjunctive form, too. One possible interpretation of this is that with *para que*, the IntHSs have developed a variable grammar in which subjunctive mood items are *preferred*, but both subjunctive, as well as underspecified indicative forms are possible

para que complements. Following this same line of thought, we can say that the AdvHSs, who were 68.0% accurate in non-presuppositional contexts, may have developed a variable polarity mood grammar. Though the AdvHSs *prefer* subjunctive in these contexts, the fact that they prefer underspecified indicative items 32.0% of the time suggests that their grammars variably entertain both mood forms in this context.

In Chapter 2, it was suggested that when HSs (and bilinguals more generally) fail to access functional features during production, they will produce underspecified, default forms, e.g., indicative mood morphology in expected subjunctive contexts. This morphological outcome is predicted, as pointed out by McCarthy (2008), if we assume the framework of Distributed Morphology (Halle & Marantz, 1993), "which allows for the underspecification of vocabulary items relative to syntactic feature specifications" (McCarthy, 2008: p. 467). What is *not* predicted, however, is that HSs will produce "overspecified" forms, that is to say, morphological items which are "more specified" featurally than the feature specifications dictated by the syntax. Nonetheless, it must be acknowledged that HSs in the study exhibit considerable evidence of this unexpected pattern in both productive and receptive experimental tasks.

Not only do the HSs "overproduce" subjunctive mood in the expected indicative contexts of the CEPT, they also tend to "overaccept" subjunctive mood in the context of presupposition in the CAT (AdvHSs: 56.5%; IntHSs: 58.6%) and "over-prefer" subjunctive mood in the context of presupposition in the MPT (AdvHSs: 14.4%; IntHSs: 16.5%). Though these findings complicate the underspecification story argued for thus far, it is possible that they are the result of instability in HSs' lexical, rather than syntactic, knowledge. If an IntHS produces the "overspecified" subjunctive mood form *observe*$_{SUBJ}$

with the complementizer *porque*, for example, it may be because in his mental grammar *observe* is classified as an unmarked, default form. Though difficult to validate in this project, it is worth noting that this speculative possibility would, nonetheless, be consistent with the lexical frequency effects shown in Chapter 5 and recapped in Section 6.3.2.

In summary, HSs exhibited *variability* in their production, acceptance and preference of subjunctive mood forms in Spanish. To answer the second part of RQ #1, therefore, we can state that HSs' subjunctive knowledge is *variable* and not categorical. Not only do HSs sometimes produce indicative mood forms when subjunctive is expected, they also accept, and even sometimes prefer, infelicitous indicative items, suggesting that their mood variability is not limited to tasks which tap into productive knowledge.

6.2. Between-group factors

Thus far, we have seen that HSs' knowledge of subjunctive mood is both *divergent* and *variable*. Nonetheless, the nature of these patterns remains unclear. What factors make HSs more or less likely to exhibit divergence and/or variability? RQ #2, copied below from Chapter 4, targets the between-group perspective of this broader question.

> 2. What is the role of between-group factors on HSs' knowledge of subjunctive mood?

In Section 6.2., I will revisit the data from Chapter 5 to review the effects of the two between-group factors that were examined in the present study: (1) Spanish proficiency (Section 6.2.1) and (2) AofA Eng (Section 6.2.2). In order to provide a more fine-grained analysis of each variable's effect, I will supplement the group results with a

presentation of HSs' individual data. To conclude Section 6.2.1 and Section 6.2.2, respectively, I will consider whether, and to what extent, the findings are consistent with each of the three approaches to HS divergence that were presented in Chapter 2.

6.2.1. Spanish proficiency

In the present study, Spanish proficiency, as operationalized by the DELE, was an excellent predictor of HSs' divergence and variability. Following previous research (e.g., Montrul & Slabakova, 2003), the HSs in the study were divided into two proficiency groups: HSs who scored between 30 and 39 were classified as intermediate (IntHSs) while HSs who scored between 40 and 50 were classified as advanced (AdvHSs). As demonstrated in Chapter 5, there were consistent differences between these two proficiency groups across all three subjunctive experiments.

In the CEPT, the AdvHSs' odds of producing subjunctive mood were many times higher than the odds for the IntHSs with both intensional ($p < .001$, OR = 7.16) and polarity ($p < .001$, OR = 8.81) subjunctive forms. Higher proficiency HSs, not surprisingly, are much more likely to produce subjunctive mood forms, at least in expected subjunctive conditions. Interestingly, proficiency differences also predicted the extent to which HSs "overproduced" subjunctive, specifically with the complementizer *porque*. The IntHSs, despite a much lower overall probability of producing subjunctive, were marginally more likely than the AdvHSs to produce non-target subjunctive mood forms with *porque*. This unexpected "overproduction" of subjunctive mood suggests increased lexical instability, e.g., not knowing the mood inflection of certain verbs, in the minds of the IntHSs.

In the CAT, proficiency once again was an accurate predictor of HSs' sensitivity to mood in adjectival relative clauses. While both AdvHSs and IntHSs were more likely to accept indicative mood in the presupposition condition, only the AdvHSs were also more likely to accept subjunctive mood in the non-presupposition condition, thereby demonstrating target-like mood comprehension. As was also the case in the CEPT, proficiency predicted an unexpected linguistic behavior. For the IntHSs, who were over two times more likely to accept indicative than subjunctive forms in the CAT ($p < .001$, OR = 2.46), indicative mood forms appear to be approaching default status in relative clauses. The AdvHSs, on the other hand, did not show an overall indicative mood bias ($p > .9$, OR = 0.97), thereby differing qualitatively from the lower-proficiency IntHSs.

In the MPT, proficiency once again predicted both quantitative and qualitative differences between the HSs. The AdvHSs (93.4% accuracy) were significantly more accurate than the IntHSs (80.9% accuracy) with *para que*, where the IntHSs more often preferred non-target indicative complements. In the polarity subjunctive items, on the other hand, proficiency predicted qualitative differences between the two groups. The AdvHSs (68.0% accuracy), despite considerable variability, preferred subjunctive in non-presuppositional relative clauses. The IntHSs (45.0% accuracy), on the contrary, did not show an overall preference for either indicative or subjunctive mood forms, apparently treating the two morphological forms as equivalent in meaning.

We have seen, thus far, that broad proficiency divisions accurately predict both quantitative and qualitative differences between HSs in productive and receptive tests of subjunctive mood. By subdividing both the AdvHSs into smaller groups, we can see that

DELE scores also predict HS variability even *within* the advanced and intermediate proficiency ranges, respectively.

To illustrate this, I subdivided the AdvHSs and IntHSs into "high" and "low" subgroups. The "low" subgroup of AdvHSs consisted of the 11 HSs who scored between 40 and 42 on the DELE (Mean: 40.83) while the "high" subgroup consisted of the 11 HSs who scored between 43 and 45 on the DELE (Mean: 44.0). Similarly, the "low" subgroup of the IntHSs consisted of the 8 IntHSs who scored between 30 and 34 on the DELE (Mean: 32.1) while the "high" subgroup consited of the 12 IntHSs who scored between 35 and 39 (Mean: 36.59). As we can see in Figure 16 below, the "high" proficiency HS subgroups produce significantly more subjunctive mood forms than their "low" proficiency subgroup counterparts, suggesting that Spanish proficiency, as operationalized by the DELE, is predictive even *within* the broad proficiency ranges typically used.

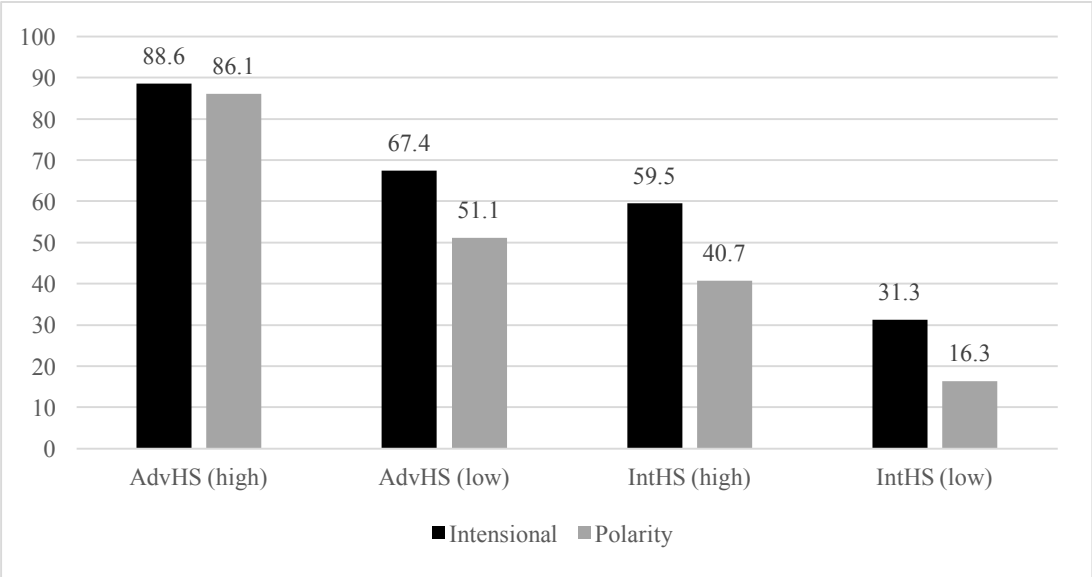

Figure 16: HSs' use of subjunctive by proficiency subgroup

It is important to note, however, that Spanish proficiency is not *fully* deterministic, as can be illustrated by a cursory look at the individual data. Participant #33, despite scoring

41 on the DELE and placing into the AdvHS group, produced only one total instance of subjunctive mood in 24 (4.1%) experimental items from the CEPT where subjunctive was expected. Scoring highly on the DELE, therefore, does not guarantee that a HS will have subjunctive mood in his/her repertoire. Participant #32, on the other hand, scored 36 on the DELE, placing into the IntHS group, and yet produced subjunctive mood in 21 out of 24 (87.5%) total experimental items where subjunctive was expected. Scoring below the advanced proficiency range, therefore, does not imply that a given speaker will lack consistent productive knowledge of subjunctive mood.

What causes Participant #33 to "underperform" with respect to her proficiency score and Participant #32 to "overperform" with respect to his? One possibility is that other between-group factors (e.g., HL use) may conspire to override the strength of the proficiency effect. Participant #33 was one of only 3 AdvHSs in the study who reported using no Spanish at Rutgers (either with friends or at work). In addition to not using Spanish at Rutgers, Participant #33 also reports that she does not use Spanish at all (0%) with any of her three younger siblings, suggesting that her family has shifted strongly away from using Spanish in the household. Participant #32, on the other hand, reports using exclusively Spanish with both parents. As an only child, his parents' strict Spanish-only policy ensured that he would have ample opportunities to use his HL throughout childhood. It is also worth noting that Participant #32's intermediate proficiency may be "offset" by his late AofA Eng, as he began acquiring English at age 5. (See Section 6.2.2.)

It is worth considering, at this point, why proficiency is such a clear predictor of HSs' subjunctive knowledge in both productive and receptive tasks, as has also been found in other studies of HSs' subjunctive knowledge (e.g., Montrul, 2009; Montrul & Perpiñán,

2011; Perez-Cortes, 2016). What makes proficiency, and more specifically, the DELE proficiency examination, such a reliable predictor of HSs' linguistic behavior?

It is important to note that the DELE is largely an examination of lexical, rather than grammatical (e.g., morphosyntactic), knowledge. The first 30 questions of the DELE test participants' knowledge of vocabulary ranging from basic, highly frequent words (e.g., *¿Qué tal?* ('What's up')) to high register, highly infrequent words (e.g., *naufragio* ('shipwreck')). Only the final 20 questions tap into participants' knowledge of grammatical knowledge, ranging from lexical aspects of the grammar, such as subcategorization of verbs, to inflectional morphology. Of these twenty items, it must be mentioned, two evaluate participants' knowledge of subjunctive mood morphology.

Given that so much of the test is evaluating lexical knowledge, it's possible that the DELE serves as an unintentional proxy for HSs' experience and use of the language. It is probably the case that HSs who know lower register words such as *naufragio* are more likely to have extensive use of and exposure to Spanish. Therefore, speakers scoring more highly on the DELE are more likely to perform in a more target-like manner with subjunctive mood, despite the fact that there is no obvious connection between recognizing the Spanish word for *shipwreck* and utilizing subjunctive mood morphology. Consistent with this speculative claim is work by Polinsky (2006), who has found that Russian HSs' maintenance of nominal case can be predicted quite accurately by the size of their receptive vocabulary in Russian. We will return to further discuss the role of lexical knowledge in HSs' variability in Section 6.3 of the present chapter.

To conclude Section 6.2.1., we will now briefly consider how this proficiency effect might be explained by each of the three approaches to HS divergence/variability.

The Input Quality approach (e.g., Pires & Rothman, 2009) could presumably explain the strong proficiency effect by arguing that HSs with higher DELE proficiency are exposed to higher quality Spanish input, in one of two potential senses. On one hand, HSs who score higher on the DELE may be less likely to be exposed to Spanish that has been significantly impacted or changed by exposure to English. In the case of subjunctive, this could mean that IntHSs are more likely to hear, in the input that they receive, instances of indicative mood produced when subjunctive is expected.

It may be the case, for example, that the IntHSs in the present study received subjunctive mood input that was more variable or qualitatively different than the subjunctive mood input received by the AdvHSs. One way that this could happen, of course, would be if more of the IntHSs had older siblings, and, consequently, received more exposure to the Spanish of other early-childhood bilinguals. Consistent with this speculative explanation is the fact that in the present study, more of the IntHSs (57.8%) had older siblings (who are, themselves, HSs) than the AdvHSs (42.8%). Though the two proficiency groups may have differed in the input quality that they received from other HSs, it is still highly unlikely that they were received different input quality from first-generation immigrants, given that the first-generation immigrants (SDCs) in the present study produced almost zero instances of indicative mood in expected subjunctive contexts.

The second possible input quality explanation is that the HSs who score in the advanced range are more likely to exhibit subjunctive knowledge because they are more likely to have received exposure to a greater breadth of styles and registers in Spanish, e.g., via formal schooling (Kupisch & Rothman, *forthcoming*; Bayram et al, 2017). Though this is a very reasonable possibility, neither intensional subjunctive with *para que* nor polarity

subjunctive in adjectival relative clauses is limited, by any means, to formal register Spanish, meaning that even HSs with less breadth of Spanish exposure are nearly certain to have heard both of these forms in informal contexts.

The Representational Differences approach (e.g., Montrul, 2002, 2008, *inter alia*) assumes that HSs who exhibit divergence with a given property have different underlying representations of that property. Since both HS groups diverge from the SDCs, this approach would presumably have to assume that both groups are representationally distinct from the SDCs with respect to their knowledge of subjunctive mood. Following this logic, the RD approach would also have to assume that AdvHSs, who perform more accurately across all three tasks, have different, and presumably more target-like, representations of subjunctive mood than the IntHSs. What remains unclear, however, is *how* exactly the AdvHSs' subjunctive knowledge is underlyingly different from that of the lower-proficiency IntHSs. (Do the IntHSs have a "more incomplete" knowledge of the underlying syntactic features?) Crucially, there is nothing inherent in the RD approach that connects HS differences in grammatical knowledge to lexical knowledge more generally. Without further elaboration of the grammatical underpinnings which drive the observed proficiency differences, the RD approach is only able to "restate the problem," e.g., by asserting that AdvHSs perform more accurately because they have a more accurate system.

The Activation/Lexicalist approach (Putnam & Sánchez, 2013) is most consistent with the proficiency distinctions presented thus far. If proficiency, as measured by the DELE, is largely lexical, and HSs' subjunctive variability is at least partially triggered by gaps in lexical, rather than morphosyntactic knowledge, then it makes sense that HSs who

score higher on a lexical proficiency test would be more likely to produce and comprehend subjunctive mood morphology.

To conclude Section 6.2.1., Spanish proficiency is a highly predictive, though not totally deterministic, factor in shaping HSs' demonstrated knowledge of subjunctive mood morphology. In the next section, we will consider how AofA Eng affects HSs' likelihood of producing and comprehending subjunctive mood in Spanish.

6.2.2. Age of acquisition of English (AofA Eng)

A second between-group variable which affected participants' demonstrated subjunctive knowledge was AofA Eng. In the present study, participants were divided into three broad groups based on when they began acquiring English. The HSs began acquiring English between birth and age 6, the LCIs began acquiring English between age 8 and age 12, and the SDCs began acquiring English at age 13 or later. In the paragraphs that follow, I will summarize the effect of AofA Eng by comparing the AdvHSs to the LCIs and the SDCs. The IntHSs will be excluded from this initial discussion because their differences from the LCIs and SDCs could be attributable to lower proficiency *or* AofA Eng.

In the CEPT, AofA Eng played a very strong role in predicting subjunctive variability. Both the LCIs and the SDCs, respectively, produced subjunctive categorically in expected subjunctive contexts. The AdvHSs, on the other hand, performed much more variably, producing occasional indicative mood forms in expected subjunctive contexts. On the basis of these three groups' subjunctive production, it appears that bilingual native Spanish speakers are way more likely to exhibit subjunctive variability if they are exposed

to English between birth and age 6. If they begin learning English at age 8 or later, however, they are highly unlikely to show mood variability.

In the CAT, AofA Eng played a less prominent role, given that the SDCs, LCIs, and AdvHSs all showed target-like sensitivity to mood morphology in adjectival relative clauses. There was, however, one subtle difference between the groups that may be a consequence of AofA Eng. In general, the AdvHSs (overall acceptance: 79.8%) were more likely than the LCIs (overall acceptance: 59.4%) and the SDCs (overall acceptance: 66.6%) to accept items in the task, suggesting that earlier AofA Eng may lead bilinguals to have less determinant judgments than bilinguals with earlier AofA Eng (e.g., Polinsky, 2006).

Similarly, in the MPT, AofA Eng played a strong predictive role, as the AdvHSs were less likely to respond accurately as compared to both the LCIs and SDCs, respectively. Based on this group difference, it seems that native Spanish speakers with later acquisition of English appear to have more consistent mood preferences with both intensional and polarity subjunctive selection.

Broadly, we can conclude that later exposure to English increases native Spanish speakers' demonstrated productive and receptive knowledge of subjunctive mood. It would not be an overstatement, based on the data from the LCIs and SDCs, to suggest that native Spanish speakers who do not begin learning English until age 8 or later are almost certain to categorically produce the intensional and polarity subjunctive mood forms tested in the present study. (For one rare exception to this generalization, see Montrul's (2013) study of a Guatemalan adoptee who experienced a dramatic reduction in input at age 10.)

If we are to understand HSs' mood variability, as opposed to bilingual mood variability more generally, it is critical that we supplement this first analysis with a study

of the effect of AofA Eng on *HSs'* subjunctive mood variability. When we control for the potentially confounding effect of proficiency, are HSs with earlier AofA Eng more likely to exhibit variability in their knowledge of subjunctive?

To make this comparison, I divided both the AdvHSs and the IntHSs into three, roughly equal sub-groups (Late AofA, Mid AofA, and Early AofA) based on their self-reported AofA Eng. Because the HSs in each proficiency group were not evenly distributed with respect to AofA Eng, the cut-off ages for dividing participants are not the same for the AdvHSs and IntHSs, as illustrated in Tables 48 and 49 below.

Table 48.
AdvHSs: AofA Eng subgroups

	AofA Eng Range	**AofA Eng Average**	**Average DELE**
Late AdvHSs (n=6)	5-6 years	5.33 years	42.50
Mid AdvHSs (n=8)	3.5-4 years	3.81 years	42.50
Early AdvHSs (n=8)	0-3 years	1.38 years	42.25
Total (n=22)	0-6 years	3.34 years	42.41

Table 49.
IntHSs: AofA Eng subgroups

	AofA Eng Range	**AofA Eng Average**	**Average DELE**
Late IntHSs (n=8)	5-6 years	5.25 years	35.25
Mid IntHSs (n=5)	2-4 years	3.20 years	35.20
Early IntHSs (n=7)	0-1 years	0.43 years	34.00
Total (n=20)	0-6 years	3.34 years	34.80

To gauge the effect of AofA Eng on HSs' subjunctive mood knowledge, I then briefly compared the experimental performance of the AdvHS and IntHS AofA Eng subgroups in the CEPT and CAT tasks, respectively. Prior to presenting the results of these comparisons, it is important to note that differences between the performance of the AofA Eng groups are almost certainly not due to differences in DELE proficiency: as shown in

Tables 48 and 49, the DELE scores of the AofA Eng subgroups are nearly identical in the case of both the AdvHSs and the IntHSs.

Figure 17 below reveals that AofA Eng has a strong effect on AdvHSs' production of subjunctive mood in the expected subjunctive conditions of the CEPT. AdvHSs with the latest AofA Eng perform nearly categorically in the production of both intensional (Mean: 95.1%) and polarity (Mean: 88.2%) subjunctive mood. AdvHSs with the earliest AofA Eng, however, perform substantially more variably with both intensional (Mean: 68.8%) and polarity (Mean: 57.9%) subjunctive mood, respectively.

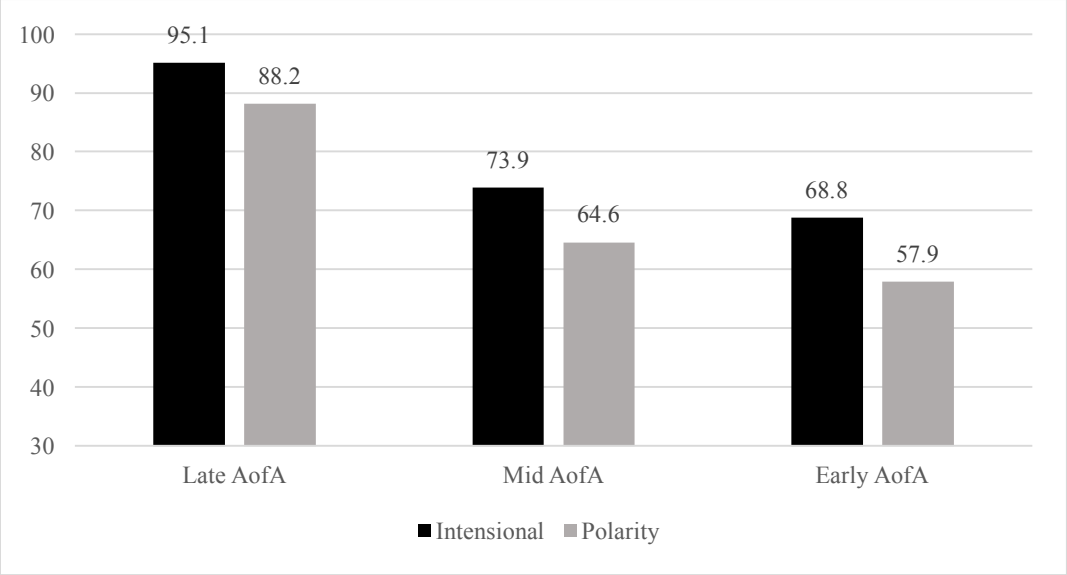

Figure 17: AdvHSs' use of subjunctive: AofA Eng x MoodSelectType

Figure 18 reveals that AofA Eng has a similar, though less pronounced effect on IntHSs' production of subjunctive mood morphology. IntHSs with the latest AofA Eng produce more intensional (Mean: 55.4%) and polarity (Mean: 32.2%) subjunctive than IntHSs with the earliest AofA Eng, who produce very little subjunctive with both intensional (Mean: 38.9%) and polarity (Mean: 24.0%) mood selection, respectively.

Interestingly, there appears to be little difference between the IntHSs with Late AofA Eng and Mid AofA Eng, perhaps due to the small sample.

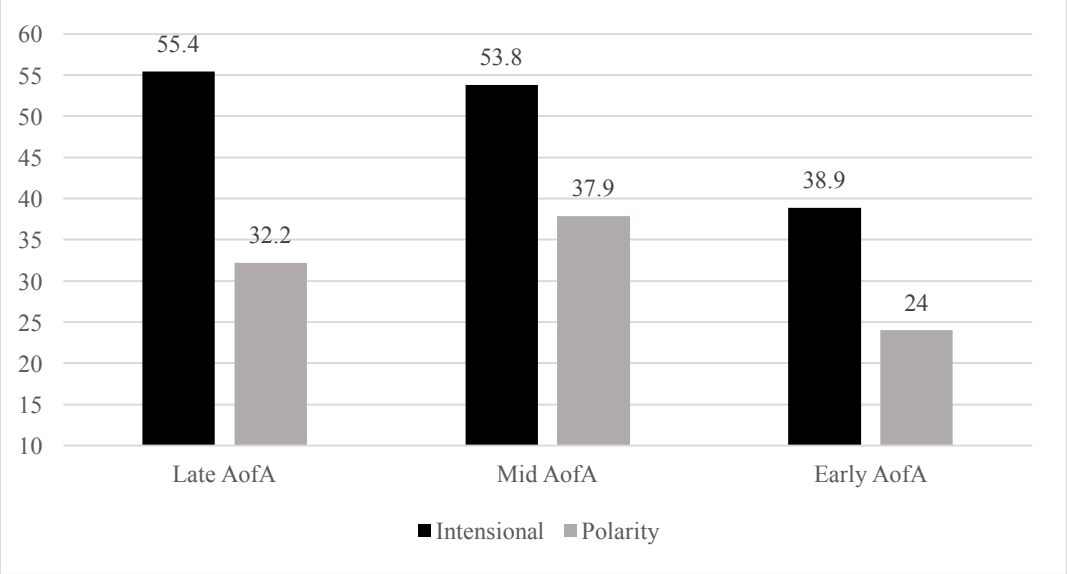

Figure 18: IntHSs' use of subjunctive: AofA Eng x MoodSelectType

The effect of AofA Eng was also apparent in the CAT, which tested HSs' mood sensitivity in adjectival relative clauses. Recall that participants in the CAT rated sentences using a five-point acceptability scale (1=sounds very good; 5 = sounds very odd), which was later converted to a binary acceptability scale (1=accept; 0 = reject). In this section, for ease of exposition, I will illustrate the AofA Eng effect on subjunctive comprehension by using participants' average ratings on the original five-point scale.

As shown in Figure 19, the AdvHSs with the latest AofA Eng make the strongest mood distinctions of the three AofA Eng groups in both the Presupposition and No Presupposition conditions. The AdvHSs with mid- or early AofA Eng make the same qualitative mood distinctions (e.g., rating indicative higher with Presupposition and subjunctive higher with No Presupposition) as the AdvHSs with the latest AofA Eng.

Nonetheless, it is apparent that these two subgroups distinguish less clearly between indicative and subjunctive mood morphology in adjectival relative clauses of the CAT.

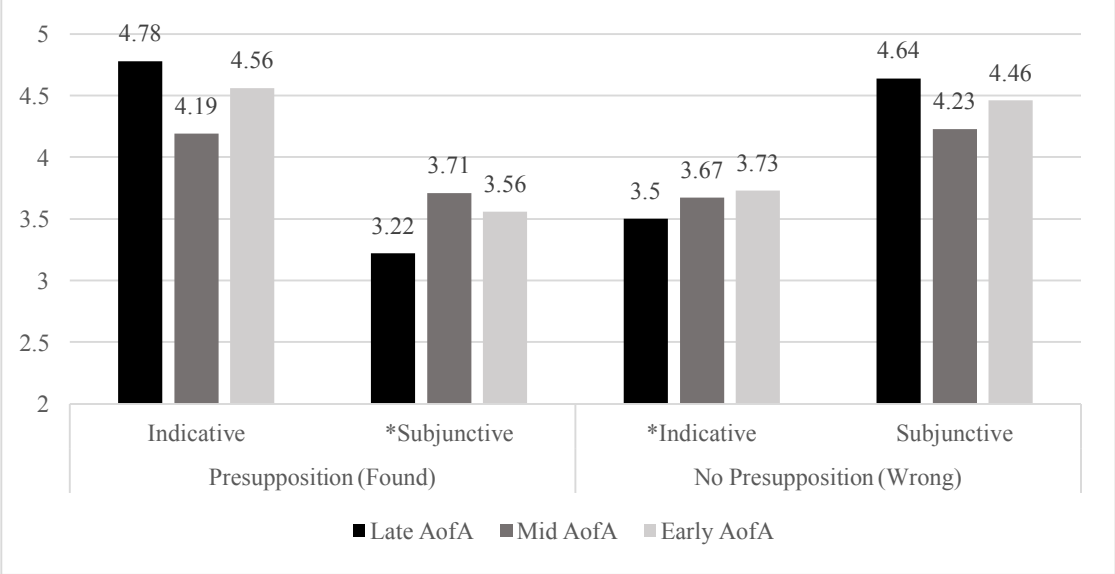

Figure 19: AdvHSs' acceptability ratings by AofA Eng, Condition

Figure 20 shows that AofA Eng has a weaker effect on IntHSs' mood sensitivity in adjectival relative clauses of the CAT, especially in the Presupposition conditions. In the No Presupposition conditions, however, AofA Eng seems to predict qualitatively different behavior from the subgroups. For the IntHSs with the latest AofA Eng, it seems that indicative (Mean: 3.87) and subjunctive (Mean: 3.85) mood morphology are equally acceptable in the absence of presupposition, perhaps because these two mood forms are in free variation. For the IntHSs with earliest AofA, however, infelicitous indicative mood items (Mean: 3.81) are more acceptable than felicitous subjunctive items (Mean: 3.43), suggesting that in these HSs' grammatical systems, indicative mood is becoming the default mood morphology in all adjectival relative clauses. Interestingly, the IntHSs in the middle AofA Eng subgroup are the only IntHSs to rate subjunctive items higher in the No

Presupposition condition. This effect, however, is driven almost exclusively by two individual HSs (HS #1 and HS #5) with high DELE scores (38/50 and 36/50).

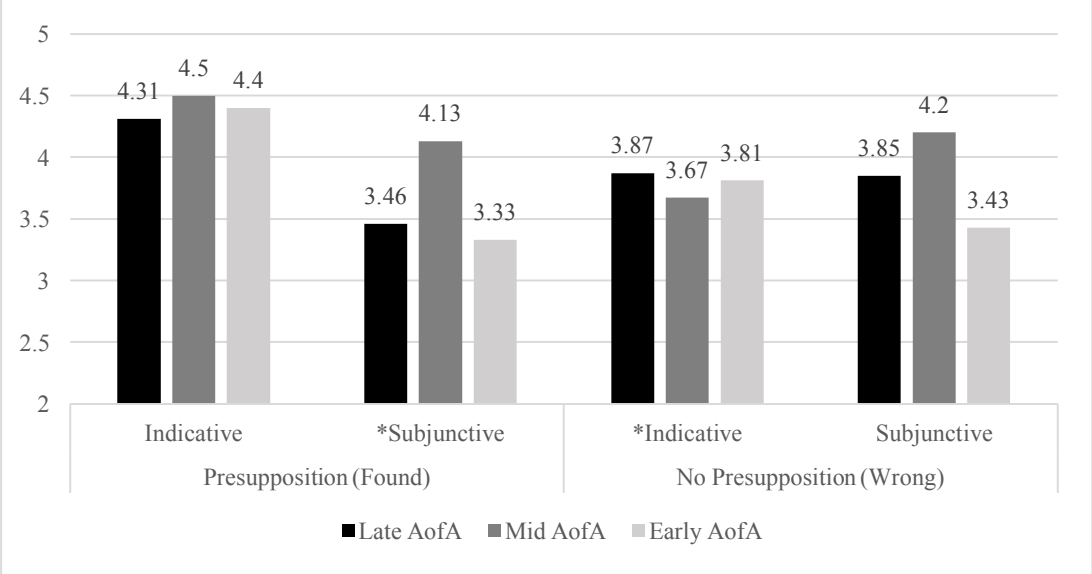

Figure 20: IntHSs' acceptability ratings by AofA Eng, Condition

The subgroup analysis presented thus far provides strong evidence that AofA Eng affects HSs' production and comprehension of subjunctive mood morphology in Spanish. The general trend, not surprisingly, is that HSs with later AofA Eng tend to show less overall variability with subjunctive mood, for reasons to be discussed later in this section.

Though the effects of AofA Eng are clearly apparent, it is important to highlight that AofA Eng, like Spanish proficiency, is not a fully deterministic predictive variable. This can be most easily illustrated by highlighting individual HSs who seem to defy expectations based on their AofA Eng. Participant #4, an AdvHS who acquired English from birth, produced subjunctive mood in 23 of 24 (95.8%) expected subjunctive contexts in the CEPT, showing that HSs with early AofA Eng can, nonetheless, perform (nearly) categorically with subjunctive mood. Participant #31, an IntHS who produced subjunctive mood in just 3 of 21 (14.3%) expected contexts, illustrates the opposite point, namely that

acquiring English later (age 5, in this participant's case) does not guarantee more consistent subjunctive production. While AofA Eng clearly shapes HSs' likelihood of producing and understanding subjunctive, it must be the case that other factors (e.g., frequency of language use: Perez-Cortes, 2016) can mitigate this strong effect and lead HSs to defy expectations with respect to their exhibited grammatical knowledge.

Participant #4 may "outperform" his early AofA Eng due to high Spanish usage with parents (90% reported Spanish usage with Mom; 99% reported Spanish usage with Dad), at work (50% reported Spanish usage), and in his broader community. In addition to traveling back to his parents' home country every 2-3 years, Participant #4 also reports using only Spanish at church and mostly Spanish with neighbors.

Participant #31, on the other hand may "underperform" her early AofA Eng as a consequence of relatively low Spanish language usage with parents (25% reported Spanish usage with Mom; 0% reported Spanish usage with Dad), who typically constitute HSs' primary source of primary linguistic input in the HL. Participant #31's reduced Spanish usage at home may also be a consequence of the fact that her parents themselves are 2nd generation HSs, unlike all other HSs' parents in the present study.

As in previous HL research (e.g., Lee, 2011; Montrul, 2002; Pascual y Cabo & Gómez Soler, 2015, Silva-Corvalán, 2014; although see Montrul & Sánchez-Walker, 2013), AofA Eng played a substantial role in predicting the extent to which HSs exhibit grammatical variability. The consistency of this finding, both in previous research and in the present study, raises an important question. Why do HSs with earlier AofA Eng so consistently exhibit increased grammatical variability?

One explanation to consider is that HSs with earlier AofA Eng receive significantly less overall Spanish input than HSs with later AofA Eng and, consequently, have less input with which to develop categorical grammatical knowledge. While it is likely that increased Spanish input quantity helps HSs acquire the HL, it is unclear how exactly increased input would lead HSs to develop more categorical linguistic knowledge. Without knowing how much input is needed for HSs to categorically acquire a property (Putnam & Sánchez, 2013), this line of inquiry becomes exceedingly difficult to investigate.

A second possibility is that an earlier AofA Eng leads to more (and earlier) opportunities for cross-linguistic influence (Müller & Hulk, 2001) in the minds of HSs. With some properties, e.g., bare nominals (Montrul & Ionin, 2010), this cross-linguistic influence can be clearly identified as having come from the societally-dominant language. With other properties, however, the influence of the societally-dominant language may be less easily apparent, e.g., when the dominant-language does not have an "equivalent" property to transfer. In these cases, the early influence from the dominant language may be less direct. It is conceivable, for example, that HSs who begin using English early may have increased difficulty producing certain functional features in the HL due to reduced activation of the HL for production (Putnam & Sánchez, 2013).

To conclude Section 6.2.2., we will now briefly consider how the AofA Eng effect ilustrated in the present study might be explained by each of the three approaches to HS divergence/variability.

It is conceivable that the Input Quality approach (Pires & Rothman, 2009) could account for some of the differences observed between HSs with "early" and "late" AofA Eng. In the previous section, we saw that HSs' Spanish proficiency-level was shown to

correlate with their sibling status, such that IntHSs were more likely than AdvHSs to have received early Spanish exposure (input of different "quality") from older siblings. Consequently, it was argued that the proficiency effects observed in the present study may be somewhat attributable to differences in the input quality received by HSs of different proficiency levels. Similarly in the case of AofA Eng, it may be that HSs in the "earlier" AofA Eng groups are, like the IntHSs in our proficiency analysis, more likely to have received early Spanish exposure (input of different "quality") from older siblings. This possibility is not at all farfetched given that the HSs' earlier AofA Eng is likely the direct result of the presence of older siblings who brought English from school into the home.

It is not likely, however, that the AofA Eng effects illustrated here are the result of differences in the input quality that HSs have received from first-generation speakers. As argued in Section 6.2.1., the first-generation SDCs showed essentially no variability at all in the production of subjunctive mood, implying that different types of HSs are unlikely to be exposed to substantial differences in input quality from first-generation speakers.

Presumably, the Representational Differences approach would have to explain the AofA Eng effect by assuming that "early" HSs have different (and more "incomplete") underlying grammatical representations than their "late" HS counterparts. From the perspective of this approach, it is not clear, however, *how* and *why* early exposure to English would lead HSs to develop different underlying grammatical representations.

The Activation/Lexicalist approach could explain the strong AofA Eng effect by positing developmental differences in the stability of the HL lexicon. If we assume, following Putnam & Sánchez (2013), that language acquisition involves (i) learning functional features and then (ii) mapping those features onto language-specific lexical

items, then "early" HSs' increased variability could be traceable to instability in their mapping of functional features to lexical items. In other words, learning English early may affect the way in which these Spanish HSs map functional features to particular lexical items. I will discuss this possibility in more detail throughout Section 6.3.

6.2.3. Summary of between-group factors

Thus far in Section 6.2., we have seen that both Spanish proficiency and AofA Eng have a strong effect on HSs' variability with subjunctive mood, though this effect is not completely deterministic. The best way to conceptualize these between-group effects, therefore, is to think in terms of shifting probabilities. All else equal, HSs with higher DELE scores are much less likely to exhibit variability in their subjunctive knowledge. In the same way, HSs who begin acquiring English later in childhood are also less likely to exhibit variability in subjunctive knowledge. Nonetheless, as we have seen, not all HSs with these specific characteristics come to demonstrate strong and target-like knowledge of subjunctive mood. In some cases, additional factors such as reported usage of the HL can supersede the strong effects of proficiency and AofA Eng.

6.3. Within-group factors

In Section 6.2., we reviewed two *between-group* factors that make some HSs more likely to exhibit variability with subjunctive mood. While critically important, this *between-group* perspective does not, and cannot, address the source of variability apparent in the mind of *individual* HSs. In order to fill this critical gap, therefore, the present section focuses on the *within-group* perspective of HS variability, as expressed in RQ #3 below.

3. What is the role of within-group factors on HSs' sensitivity to subjunctive mood?

The goal of Section 6.3. is to understand the linguistic factors which make some HSs more or less likely to exhibit subjunctive mood variability with any given experimental item. Each of the three factors that will be considered in Section 6.3 directly involve the HL lexicon, which is hypothesized to be a primary locus of HSs' grammatical variability.

6.3.1. Mood selection type

The present study explores HSs' knowledge of two types of subjunctive mood morphology: intensional subjunctive and polarity subjunctive. As illustrated in Chapter 3, these two types of subjunctive mood morphology are different in a few key ways, despite sharing the same surface form. (In other words, the subjunctive inflection of *tener* is *tenga*, regardless of whether the verb appears in an intensional or polarity subjunctive structure.) In the present section, we will focus on the way in which these two types of subjunctive are stored differently in the HL lexicon.

Intensional subjunctive is triggered by individual lexical items such as the complementizer *para que*. Because *para que* obligatorily selects for a subjunctive mood compelement, I argued in Chapter 3 that Spanish-speakers likely store *para que* alongside its complementary structure as a treelet (in the sense of Jackendoff, 2007) in their Spanish lexicon. If this is true, producing subjunctive mood morphology with *para que* should be relatively "easier" given that accessing and activating the complementizer *para que* should automatically activate its complementary structure, including the uninterpretable *u*W feature which triggers subjunctive mood morphology.

Polarity subjunctive, on the other hand, is triggered by context, rather than preceding lexical items. Recall that in adjectival relative clauses, subjunctive is used in the absence of contextual presupposition while indicative is used in the presence of contextual presupposition. As discussed in Chapter 3, the fact that polarity subjunctive is triggered by context, rather than lexical items, has significant implications for its storage in the Spanish lexicon. Specifically, it may be the case that polarity subjunctive forms require "assembly" in a way that is not true for intensional forms.

As demonstrated in Chapter 5, HSs' subjunctive mood variability was strongly conditioned by subjunctive mood type, both in productive (CEPT) and receptive (MPT) experimental tasks. In the CEPT, both the AdvHSs and the IntHSs were significantly more likely to produce subjunctive with intensional, rather than polarity mood selection. The AdvHSs' predicted probability of subjunctive production was 86.6% with intensional subjunctive compared to just 74.3% with polarity subjunctive, a difference that was significant at the .05 level. The IntHSs, despite lower overall probabilities of subjunctive production, demonstrated an even stronger effect of subjunctive mood type. The IntHSs' predicted probability of subjunctive production was 47.4% with intensional subjunctive and just 24.7% with polarity subjunctive, a difference that was significant at the .01 level. Critically, the HSs' higher likelihood of producing intensional subjunctive is not attributable to the lexical frequency of the verbs that appeared in the experimental task: as outlined in Chapter 4, the average frequency ranking of verbs in the intensional subjunctive items was not different from the ranking of the polarity subjunctive items.

Strengthening these group-level trends is an analysis of the individual production data. Of the 42 HSs in the study, only 7 (16.7%) produced a higher proportion of polarity

subjunctive than intensional subjunctive. Tellingly, only one of these 7 HSs had a proportion of polarity subjunctive production that was more than 10% higher than their proportion of intensional subjunctive production. (As a point of comparison, it is worth noting that there were *17 HSs* whose proportion of intensional subjunctive production was 10% higher than their proportion of polarity subjunctive production.) Together, these trends strongly support the hypothesis that intensional subjunctive is easier than polarity subjunctive for HSs to access and produce in Spanish.

The advantage for intensional subjunctive forms also extends to receptive experimental tasks. In the MPT, which was the only receptive task that examined both intensional and polarity forms, both the AdvHSs and the IntHSs were significantly more accurate with intensional subjunctive mood. The AdvHSs' predicted probability of responding accurately was 93.4% with intensional subjunctive compared to just 68.0% with polarity subjunctive, a difference that was significant at the .01 level. Similarly, the IntHSs' predicted probability of responding accurately was 80.9% with intensional subjunctive and 45.0% with polarity subjunctive, a difference that was also significant at the .01 level. As in the CEPT, the individual data from the MPT supplements and even strengthens the evidence that intensional subjunctive is "easier" for HSs. Out of the 42 HSs in the present study who completed the MPT, only 2 (4.8%) responded more accurately in the polarity subjunctive condition than in the intensional subjunctive condition.

To conclude Section 6.3.1., we will now briefly consider how the mood selection effect ilustrated thus far might be explained by each of the three approaches to HS divergence/variability.

The Input Quality approach (e.g., Pires & Rothman, 2009) explains that HS divergence is at least *partially* the result of exposure to qualitatively different input in the HL. How, from such a perspective, might we account for HSs' increased variability and divergence with polarity, as opposed to intensional, subjunctive mood? The obvious answer would be to suggest that HSs receive more variable input quality with polarity subjunctive in adjectival relative clauses than intensional subjunctive with *para que*. In the present study, however, there is only weak support for this possibility, at least when we consider the HL input provided to HSs by first-generation immigrants.

The Spanish-dominant controls in the present study produced both intensional subjunctive (226/226: 100%) and polarity subjunctive (227/230: 98.7%) in nearly every single expected context. Though the SDCs were slightly more variable in their production of polarity subjunctive, it is unlikely that this small difference alone would account for the substantial differences in HSs' relative command of these two types of subjunctive forms. That said, the results of the first-generation immigrants in the present study contrast with the results presented by Bookhamer (2013), who found that first-generation Spanish speakers in the US produced intensional subjunctive with *para que* 100% of the time and polarity subjunctive in adjectival relative clauses just 84.1% of the time. If Bookhamer's results are a more faithful reflection of the type of input that HSs receive from first-generation speakers, then it may indeed be the case that input quality differences play a fundamental in HSs' higher variability with polarity subjunctive mood.

It is also possible, of course, that HSs' increased variability with polarity subjunctive is the result of having received more variable input with polarity subjunctive from *other HSs*, rather than first-generation immigrants. While the HS data in the present

study is consistent with this possibility, it is important to note that such an explanation would still fail to account for why those other HSs (e.g., input providers) produce polarity subjunctive mood more variably in the first place (Putnam & Sánchez, 2013).

In light of the HSs' greater variability with polarity subjunctive, the Representational Differences approach (Montrul, 2002, 2008, *inter alia*) would seemingly have to argue that HSs have a less target-like representation of polarity subjunctive mood than intensional subjunctive mood. It's not clear, however, what exactly this might mean, especially for the AdvHSs, who produce *both* intensional and polarity subjunctive mood forms well over 50% of the time and, consequently, appear to have strong knowledge of the features that underlie each of these forms. What would cause HSs with systematic knowledge of both subjunctive types to exhibit more variability with polarity forms?

One possibility, presented in Montrul's (2009) study of HSs' verbal morphology, is that earlier acquired morphology (e.g., aspectual morphology) has more of an opportunity to stabilize than later acquired morphology (e.g., mood morphology). If we translate this same concept to different types of subjunctive mood forms, it's possible to argue that HSs' knowledge of intensional mood morphology, which is earlier acquired, has more time to stabilize than their knowledge of polarity mood morphology. This argument, though clearly intuitive, requires further and more specific grammatical elaboration. From a representational standpoint, how, and at what level of the HS grammar, does earlier exposure to the societal language lead to more variable or unstable representations?

The Activation/Lexicalist approach (Putnam & Sánchez, 2013) appears to have the best opportunity of accounting for the differences in HSs' variability with polarity and intensional subjunctive forms, respectively. By separating the acquisition of functional

features from their mapping/access, this approach acknowledges the possibility that HSs could exhibit variability with a property, even when they have acquired the functional feature underlying that property. Under this approach, the AdvHSs' well-above chance performance with both subjunctive types suggests underlying knowledge of the functional features triggering each form. Their increased variability with polarity subjunctive, therefore, is not a representational problem, *per se*, but instead the result of increased difficulty accessing or mapping that featural knowledge in real time.

Lohndal & Westergaard (2016), previously discussed in Chapter 2, provide a similar argument in their analysis of gender and definiteness heritage Norwegian. The HSs in their study produce gender more accurately with definiteness markers, which are expressed via post-nominal suffixation, e.g., -*en*, than with indefiniteness markers, which are expressed as pre-nominal free morphemes. The authors attribute the HSs' higher accuracy with gender on definiteness marking to the way it is stored in, and accessed from, the lexicon. Specifically, they posit that the definiteness suffix in Norwegian may be "initially learned as a unit together with the noun" (p. 11), making it "more easily retrieved" (p. 11) by the HSs. The indefiniteness markers, however, "must be computed as part of a productive process" (p. 11), making them more vulnerable to potential inconsistency in on-line production. It's not farfetched to think that intensional and polarity subjunctive are analogous in heritage Spanish: where intensional subjunctive is stored as a unit and therefore easily retrieved, polarity subjunctive must be computed as a productive process and, consequently, exhibits more variability in production.

In addition to these three possible explanations for HSs' higher variability with polarity subjunctive mood, it is worth briefly considering a fourth possibility, namely, that

intensional subjunctive forms with *para que* are simply much more frequent in the Spanish input than polarity subjunctive mood forms in adjectival relative clauses. Though such an explanation would be very intuitive, the available (and admittedly limited) evidence of the frequency of these structures does not support this fourth and final hypothesis.

Blake (1982) and Kowal (2007) cite a 1971 study from the Universidad Nacional Autónoma de México (UNAM) which tabulates the number and type of subjunctive mood forms produced in a small corpus of adult Spanish speakers in Mexico City. In this study, intensional subjunctive with *para que* represents 16.3% of all subjunctive tokens observed, making it over twice as frequent as adjectival relative clause subjunctive (7.8%) and (weakly) supporting the possibility of a structure frequency effect. In a larger, more recent study, however, Bookhamer (2013) reports that polarity subjunctive in adjectival relative clauses (160 total tokens) appears *more often* than intensional subjunctive with *para que* (134 tokens) in the spontaneous speech of first-generation immigrants living in New York City. On the basis of these two studies, therefore, there is no clear reason to believe that HSs are more accurate with *para que* subjunctive because they hear it more often.

6.3.2. Lexical frequency

A second within-group factor influencing HSs' subjunctive variability is lexical frequency. Throughout the present study, HSs were more likely to exhibit subjunctive variability with less frequent verbs, a finding with significant implications for HL research.

In the CEPT, participants saw twelve verbs in both the intensional and polarity subjunctive conditions, respectively. On the basis of Davies' (2006) frequency dictionary, these verbs were split into two equal groups: frequent verbs and infrequent verbs. In the

intensional subjunctive condition, both groups of HSs showed significantly less variability with more frequent verbs. The AdvHSs' predicted probability of subjunctive production rose from 76.6% with lower frequency verbs (e.g., *observar*) to 94.5% with higher frequency verbs (e.g., *hacer*), a difference that was statistically significant at the .01 level. Similarly, the IntHSs' predicted probability of subjunctive production increased from 32.6% with infrequent verbs to 59.8% with frequent verbs, a difference that was significant at the .05 level. In the polarity subjunctive condition, both groups again produced more subjunctive with frequent verbs, though in neither case was this difference statistically significant. A summary of these findings is shown in Figure 21 below.

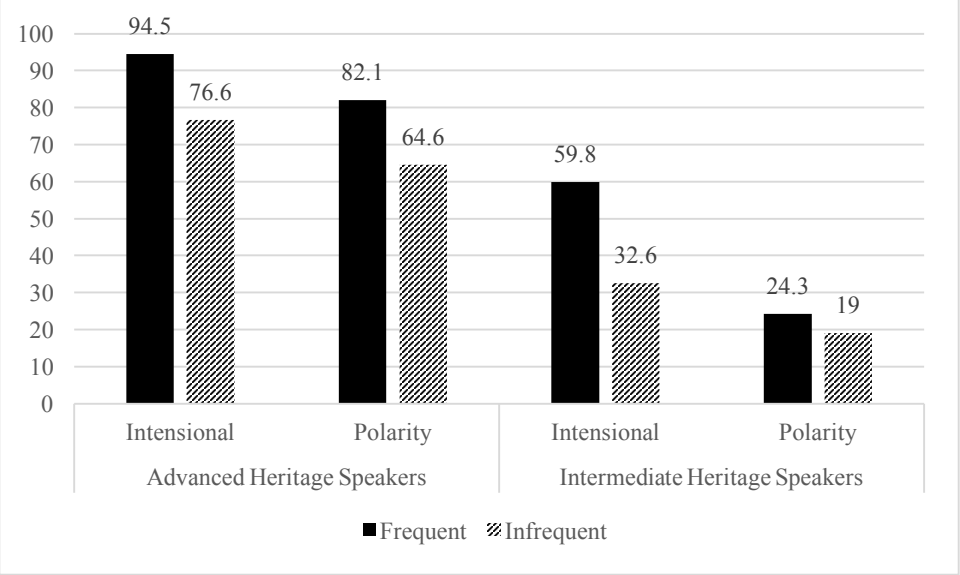

Figure 21: HSs' subjunctive use: MoodSelectType x Frequency

HSs' receptive sensitivity to mood, as measured by the CAT, was also modulated by lexical frequency, suggesting that the lexical frequency effect on HSs' variability extends beyond production. Recall that in the CAT, participants accepted or rejected sentences whose felicitousness was dependent on the mood morphology in the relative

clause. When the context provided presupposition, participants with target-like mood knowledge were expected to *accept* experimental items with indicative mood and *reject* experimental items with subjunctive mood. Alternatively, when the context *did not* provide presupposition, participants were expected to *accept* experimental items with subjunctive and *reject* experimental items with indicative.

Of the HS groups, only the AdvHSs made both distinctions, thereby demonstrating mood sensitivity in adjectival relative clauses. Curiously, though, the AdvHSs' sensitivity to mood was significantly affected by verb frequency, as shown in Figure 22 below.

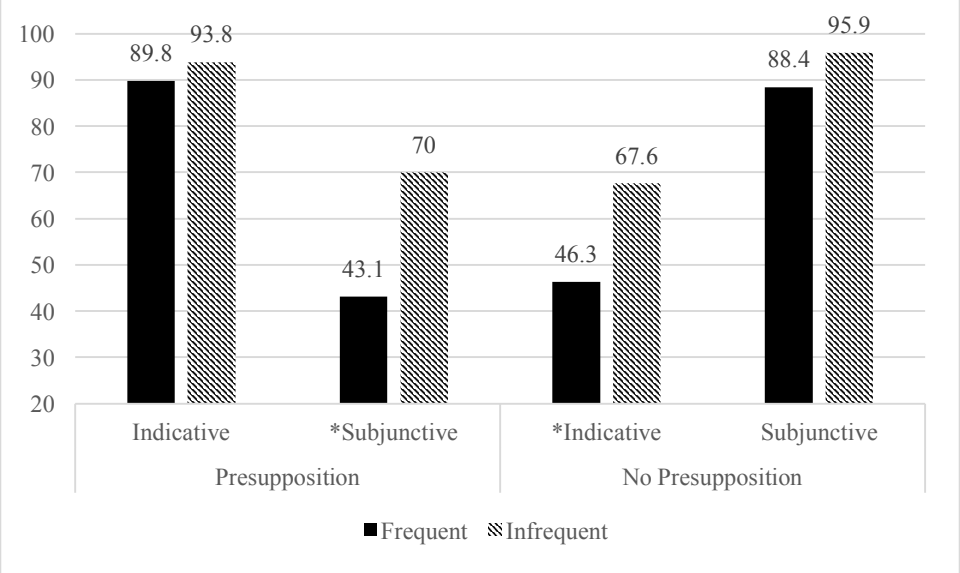

Figure 22: AdvHSs' probability of acceptance: Condition x Frequency

With the frequent verbs in the CAT (*ser, estar,* and *tener*), the AdvHSs accepted infelicitous items only 43.1% of the time in the Presupposition condition and 46.3% of the time in the No Presupposition condition, as illustrated by the black columns in Figure 22.

With infrequent verbs (*venir*, *funcionar* and *costar*), however, the AdvHSs were much less sensitive, accepting infelicitous items 70.0% of the time in the Presupposition condition and 67.6% in the No Presupposition condition, as illustrated by the striped columns[30]. The AdvHSs' clear mood sensitivity in the CAT, coupled with their tendency to "overaccept" infelicitous sentences with infrequent verbs, suggests that these HSs *know* the relationship between mood and presupposition but do not always *recognize* the mood morphology on the three infrequent verb forms in the CAT. HSs' receptive mood variability, therefore, is also significantly affected by lexical frequency.

Based on these findings, it appears that HSs, like second language learners, perform more more accurately with higher frequency lexical items. Though Dorian (1981) and Gal (1989) have presented anecdotal, non-experimental evidence that HSs of Gaelic and Hungarian are more accurate with inflectional morphology on frequent verb forms, the present study is the first experimental study, to my knowledge, to find that HS variability is modulated by lexical frequency in both production and comprehension.

The Input Quality approach does not have a clear way of explaining the lexical frequency effect presented in this section. As demonstrated in Chapter 5, both the SDCs and the LCIs produced subjunctive mood categorically with both frequent and infrequent verbs. It is unlikely, therefore, that the HSs' reduced mood sensitivity with infrequent verbs is a function of having heard variability in the input with infrequent verbs only. One could argue, however, that HSs in the study failed to consistently produce and comprehend subjunctive mood on infrequent verbs because these verbs did not appear (at all, or as often)

---

[30] The verb *venir* ('come') is clearly not an infrequent verb. In fact, it is ranked as the 105th most common word in Spanish according to Davies (2006). It is, however, less frequent than three other verbs from the CAT: *ser, estar,* and *tener*. Consequently, as explained in Chapters 4 and 5, *venir* was categorized as one of the three least frequent verbs in the task and, in the statistical analysis, analyzed accordingly.

in the limited Spanish input that they received at home[31]. In that sense, therefore, input quality could still play an explanatory role in HSs' mood insensitivity with infrequent verbs.

The Representational Differences approach lacks the flexibility to explain the lexical frequency effect presented here. If a HS's variability with subjunctive mood is caused by underlying representational differences (e.g., differences in featural acquisition), as this approach posits, then why would verb frequency have such a substantial effect on HSs' production and comprehension? I see no reason why representational difference would be modulated by lexical frequency. As discussed in Chapter 2, Montrul (2016a) has recently drawn a distinction between what she calls *acquisition*, e.g., when a HS shows measurable knowledge of a property, and *mastery* where that HS produces said property categorically (90-100%). It might be possible, employing this terminology, to suggest that the lexical frequency effect is an example of HSs acquiring, but not mastering, mood morphology, due to difficulty with less frequent verb forms. Nonetheless, this would still not provide an account for how and why frequency has such a pronounced effect.

The Lexicalist/Activation approach provides a clear story for the lexical frequency effect. Under this approach, it is possible to acquire abstract features without necessarily (a) mapping them onto all possible lexical items (e.g., verb roots) or (b) accessing those features in real time production or comprehension. Therefore, when a HS fails to produce or comprehend mood morphology on an infrequent verb form, there are at least two possible reasons why this might take place.

---

[31] It's also possible that HSs' increased variability with infrequent verbs is a direct consequence of their relative lack of Spanish literacy and Spanish formal education, as has been argued by Kupisch & Rothman (forthcoming), Bayram et al (2017), and others.

The first possibility is related to HL activation (Putnam & Sánchez, 2013). Infrequent verbs, by definition, are less frequently activated in the minds of HSs, which may increase the difficulty of accessing them (and their functional features) for production and comprehension. The frequency effect, therefore, could simply be an artifact of HSs' increased difficulty accessing features on infrequently activated verbs. The second possibility involves what I will call lexical learning. It may be the case that some HSs, despite having acquired the functional features which underlie mood, may fail to learn how those functional features are instantiated on certain lexical items. In other words, HSs may acquire mood, in an abstract sense, but not learn that it is expressed as *funcione* on the infrequent verb *funcionar* or *cueste* on the equally infrequent verb *costar*. Both possibilities involve distinguishing between abstract featural and concrete lexical knowledge, and both possibilities are almost certainly linked to relative frequencies of HL use and exposure.

The first of these two possibilities is consistent with the group frequency data, where HSs produce subjunctive mood frequently, indicating abstract featural knowledge, and yet evidence considerable variability with infrequent verbs, suggesting occasional difficulty *accessing* that knowledge on less frequently activated verbs. The second possibility, that HSs' variability is attributable to not knowing subjunctive mood inflections on certain, less frequent verbs, is also consistent with the group frequency data. Nonetheless, the strongest evidence for this second possibility would come from individual data, specifically HSs who show ceiling subjunctive knowledge with certain verbs and "floor" knowledge with other verbs. I will present such data in the following paragraphs.

Prior to exploring the highly informative individual HS data, it is important to first underscore the methodological detail that makes this particular individual data so valuable.

All of the verbs that appeared in the expected subjunctive conditions of the Mood Preference Task (MPT) also appeared in the expected subjunctive conditions of the Contextualized Elicited Production Task (CEPT). The repetition of these verbs across the CEPT and the MPT has two critically important benefits. First, testing HSs' knowledge of mood multiple times with the same verb can increase our confidence in determining whether they "know" or "do not know" the subjunctive mood inflection of these particular verbs. This benefit will prove critical as we continue the present discussion of lexical knowledge and HS variability. Second, testing HSs' knowledge of the same verbs in different experimental modalities (production and preference) sheds light on possible production/comprehension asymmetries in HSs' mood knowledge. Specifically, it will allow us to see whether, with certain verbs, HSs recognize, but do not produce, subjunctive mood morphology, a finding that would have considerable implications.

In the next few paragraphs, I will use individual data from the CEPT and MPT to illuminate the critical role of lexical frequency on HSs' mood variability. For the sake of brevity, I will limit myself to discussing a few particularly representative examples.

Consistent with the predictions of the Activation/Lexicalist approach, some AdvHSs showed categorically target-like (100%) knowledge of subjunctive mood with frequent verbs and categorically divergent (0%) knowledge of subjunctive with less frequent verbs. Participant #39, for example, both produced (CEPT) and preferred (MPT) target-like polarity subjunctive forms with the four most frequent verbs in this condition: *tener, decir, hablar* and *comprender*. However, with the least frequent verb, *aceptar*, Participant #39 neither produced nor preferred the target subjunctive form *acepte*. Participant #82, also an AdvHS, showed a very similar pattern, performing at ceiling in

production and preference with the most frequent verbs *tener* and *decir* yet failing to produce or prefer the target subjunctive form of the infrequent verb *aceptar*.

Based on their categorical accuracy in the production and preference of subjunctive with multiple Spanish verbs, both of these HSs appear to have acquired the abstract features underlying polarity subjunctive. The fact that both speakers exhibit divergent knowledge with the most infrequent verb, however, suggests that their abstract knowledge of mood is lexically limited, likely to verbs that they have used and been exposed to most often. This finding problematizes any approach to HL acquisition which categorizes HSs dichotomously, e.g., as having either *acquired* or *not acquired* abstract features such as mood. Quite on the contrary, these results seem to suggest that HSs can acquire an abstract feature and yet not apply or map it categorically to all verb stems.

In addition to supporting the possibility that HSs' mood variability is a result of not knowing the subjunctive mood inflection of less frequent verbs, the individual data also supports the possibility that HSs' mood variability is (sometimes) caused by temporary difficulties in accessing their knowledge of less frequently activated lexical items.

In Chapter 5, we saw that the IntHSs performed more accurately with intensional subjunctive in the MPT (80.9% accuracy) than in the CEPT (47.4% accuracy), a finding that highlights the importance of task modality in HL acquisition research (e.g., Montrul & Perpiñán, 2011). Further analysis of the IntHSs' performance across these two tasks suggests that this clear modality effect is shaped, to some extent, by lexical frequency.

On 28 different occasions, IntHSs exhibited "split" knowledge of intensional subjunctive mood with a given verb, e.g., responding accurately with that verb in either the CEPT or the MPT (but not both). In 25 of these 28 occurrences (89.3%), the IntHSs

responded accurately in the MPT (e.g., preferring the target subjunctive verb over a non-target indicative form) but not in the CEPT (e.g., producing non-target indicative). This robust tendency for the IntHSs to recognize, but not produce, target-subjunctive mood forms *of the same verb* is consistent with the possibility that they know the subjunctive mood instantiations of these verbs but sometimes fail to access them in spoken production.

Out of the 25 instances in which IntHSs were accurate with a single verb in the MPT but not the CEPT, 21 (84.0%) of those instances occurred with the three least frequent verbs in the intensional subjunctive condition, suggesting that HSs' difficulties accessing subjunctive mood knowledge in production are most pronounced with the verbs that they are least likely to use. From the standpoint of the Activation/Lexicalist approach, this finding is expected. Even if HSs know the subjunctive mood forms of lower frequency verbs, which they sometimes will, they should show a stronger task modality effect with these verbs as they are accessed and activated less frequently for production.

We have seen thus far that lexical frequency plays a strong role in HS variability. What remains unclear, however, is how to model this role conceptually. One possibility is to divide HSs into different "stages." Putnam & Sánchez (2013) propose four different stages through which HSs pass as they shift towards increasing dominance in the societal (as opposed to the heritage) language. In the first stage, HSs transfer or reassemble a limited set of functional features from L2 lexical items onto HL lexical items. In the second stage, this transfer increases in scale as HSs now transfer/reassemble *massive sets* of functional feeatures from the L2 to the L1. In the third stage, HSs struggle to *produce* functional features in the HL, though they may still exhibit comprehension of these features. Finally,

in the fourth stage, HSs struggle in both production *and* comprehension, likely due to substantial decreases in activation of the HL for both production and comprehension.

A visual representation of Putnam & Sánchez's 4 stages, adapted from Perez-Cortes (2016), is presented below in Figure 23.

Note that in Figure 23, I have replaced the term L1, used in Putnam & Sánchez (2013) and Perez-Cortes (2016), with HL (heritage language).

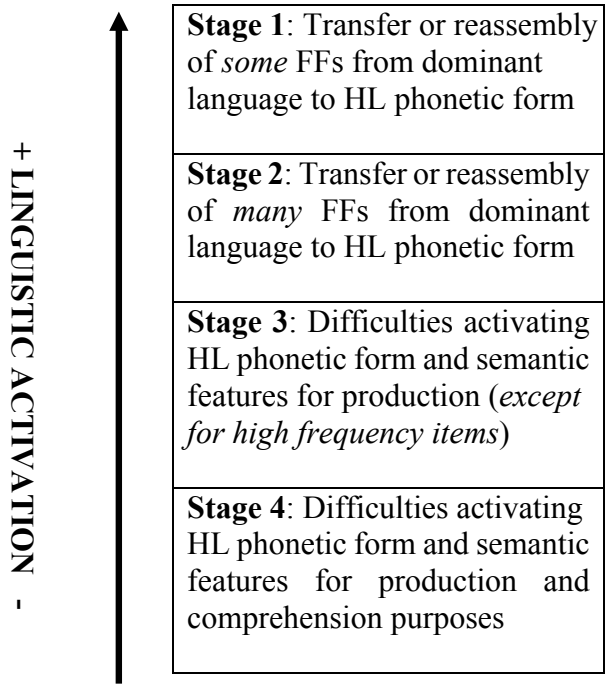

Figure 23: Four stages of HL activation: Putnam & Sánchez (2013)
(adapted from Perez-Cortes, 2016)

To highlight the role of lexical frequency in HS variability, I will borrow Putnam & Sánchez's notion of HS stages, albeit with a couple of critical differences. The first difference has to do with how variability is conceptualized. For Putnam & Sánchez, the four stages highlight differences in the extent to which functional features *transfer* or *remap* from a HS's L2 to his HL. My own conceptualization of stages, however, is broader

and not necessarily tied to the specific idea of featural transfer or reassembly[32]. The second difference has to do with the scope and generality of the proposed stages. Putnam & Sánchez take a broad view of HS variability, such that HSs in different stages differ in terms of *how many* functional features they transfer to the HL. (Note in Figure 23 that the difference between Stage 1 and Stage 2 is the *number* of FF's remapped or transferred to the HL.) In this dissertation, however, I take a narrow view, focusing exclusively on variability with subjunctive mood. Practically, this will mean that HSs in my different stages will differ from one another in terms of *how much variability* they exhibit with the features underlying mood, rather than *how many* functional features transfer into their HL.

In the present study, I divided HSs into four stages based largely, though not exclusively, on their production of subjunctive in the CEPT. Stage 1 HSs produce subjunctive 100% of the time in expected contexts and, additionally, respond with at least 80% accuracy in the MPT. Stage 2 HSs, on the other hand, exhibit variability, producing subjunctive between 50-91.7% of the time in expected contexts and responding with at least 60% accuracy in the MPT. Stage 3 HSs exhibit even stronger variability, producing subjunctive between 8.2% to 49.0% of the time in expected contexts. Finally, Stage 4 HSs produce only indicative mood forms in expected subjunctive contexts.

Now that the stages have been defined, I will illustrate the advantages of a stage-based approach by dividing up the HSs for both (a) intensional subjunctive and (b) polarity subjunctive. In dividing the HSs into four stages, I will achieve the following goals. First, I will show that HSs, even within a single proficiency range, may exhibit dramatically

[32] It is also possible, from the perspective of Putnam & Sánchez (2013), that higher English activation/usage could have a more covert impact on heritage Spanish. Instead of English features transferring directly onto Spanish lexical items, it may be the case that the *lack* of English feature specification on certain lexical items could lead certain Spanish lexical items to be unmarked featurally.

different levels of mood variability. Second, I will use lexical frequency statistics to show how the transitions from Stage 1 to Stage 2 and Stage 2 to Stage 3 entail gradually increased variability with less frequent verbs. Finally, I will use comprehension data to show that the HSs in Stages 2, 3, and 4 exhibit increasingly apparent receptive, as well as productive, variability in their knowledge of subjunctive mood.

Figure 24 shows the distribution of the AdvHSs with intensional subjunctive mood.

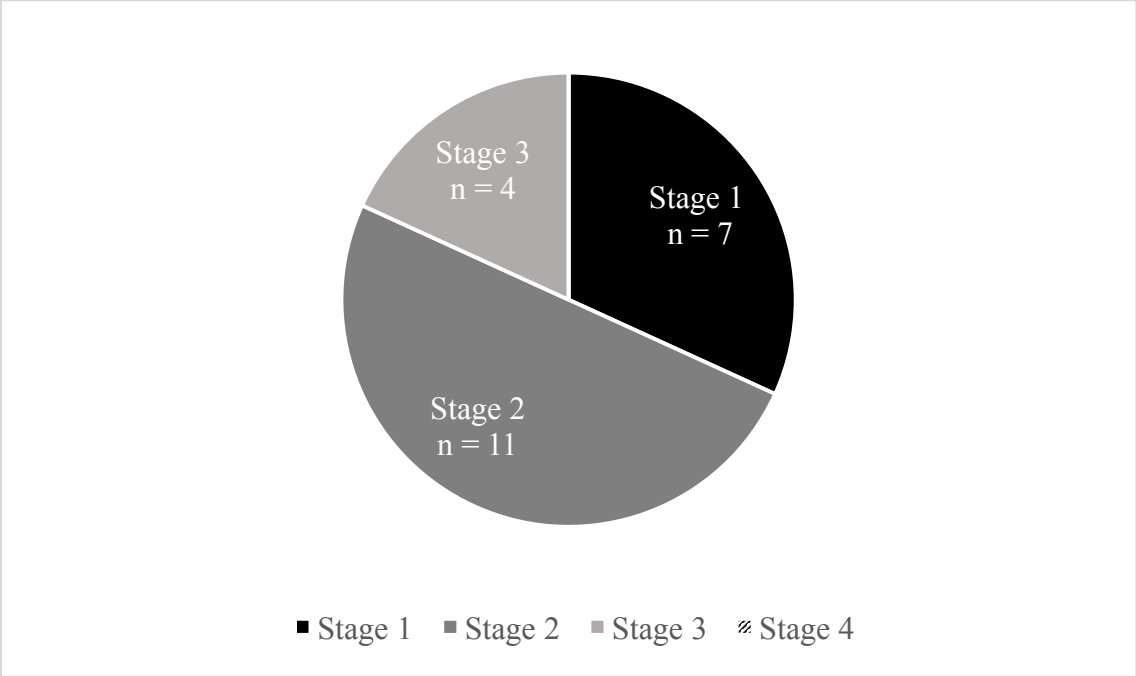

Figure 24: AdvHSs divided by stages: intensional Subjunctive

Eighteen of the 22 AdvHSs (81.8%) fall into either Stage 1 (n=7) or Stage 2 (n=11) with intensional subjunctive mood.

AdvHSs in Stage 2 retain nearly categorical (95.5%) production of subjunctive with frequent verbs, including 100% accuracy with the frequent verbs *hacer* ('do'), *parecer* ('seem'), *tomar* ('take') and *ganar* ('win'). Stage 2 AdvHSs' productive variability is most evident with infrequent verbs, which they inflect with subjunctive mood only 75.8% of the

time. Of these infrequent verbs, they are most inaccurate with *romper* ('break'; 54.5%), *observar* ('observe'; 63.6%), and *correr* ('run'; 63.6%). The key difference between Stage 1 and Stage 2 HSs is that Stage 2 HSs begin to exhibit variability with infrequent verbs.

The remaining four AdvHSs fall into Stage 3 with intensional subjunctive mood. These Stage 3 speakers produce less subjunctive mood (18.8%) than speakers in Stage 1 (100.0%) and Stage 2 (85.6%). One of the key characteristics of Stage 3 speakers is the even more noticeable effect of lexical frequency on their subjunctive mood production. The Stage 3 AdvHSs use subjunctive mood 33.3% of the time with frequent verbs and just 4.2% of the time with infrequent verbs. Clearly, these speakers' productive usage of subjunctive is mostly restricted to frequent verbs. Even with frequent verbs, though, Stage 3 speakers are far more likely to produce indicative mood than subjunctive mood.

Figure 25 shows the distribution of the IntHSs with intensional subjunctive mood.

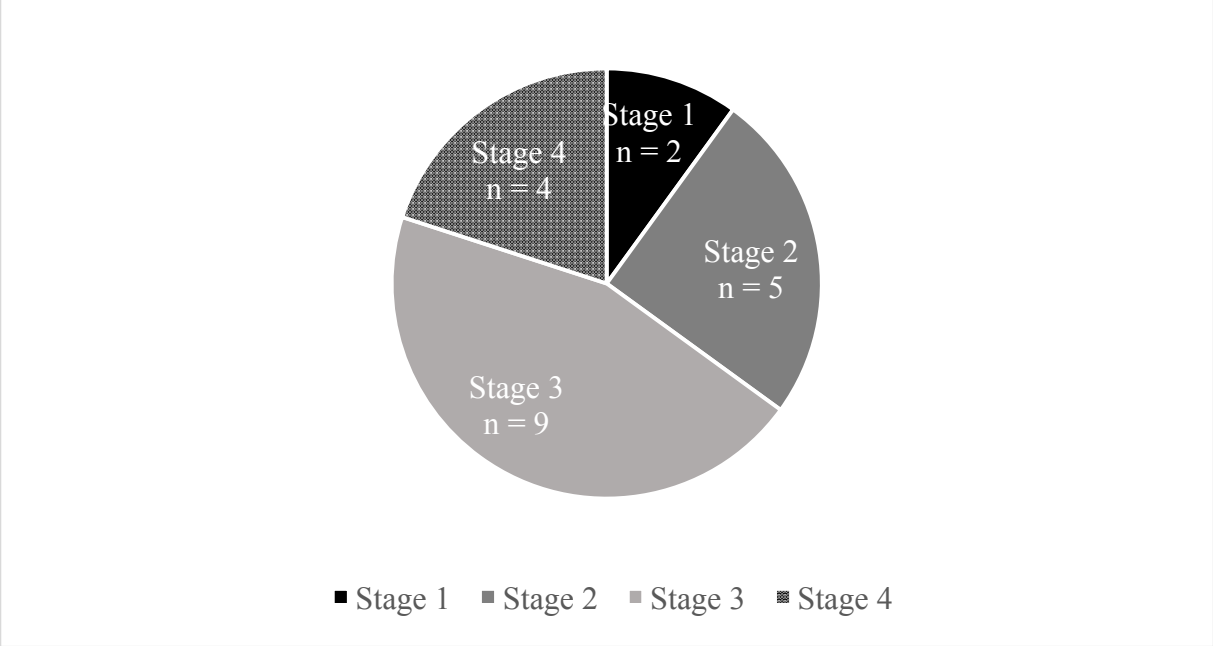

Figure 25: IntHSs divided by stages: intensional Subjunctive

Only 7 IntHSs (35%) fall into Stage 1 (n=2) or Stage 2 (n=5) with intensional subjunctive. The Stage 2 speakers, like their AdvHS counterparts, produce subjunctive almost categorically with frequent verbs (88.9%) including 100% production of subjunctive with *hacer*, *salir* ('leave'), *tomar*, and *ganar*. Their variability, as in the case of the AdvHSs, emerges with lower frequency verbs, which they produce with just 66.7% accuracy, including 40% with *observar* and 50% with *correr* ('run') and *usar* ('use').

The nine Stage 3 IntHSs produce far less subjunctive (39.8%) than the Stage 1 (100%) and Stage 2 (77.8%) HSs. Like the AdvHSs in Stage 3, the Stage 3 IntHSs exhibit strong asymmetry in production between frequent verbs, where they produce subjunctive 57.1% of the time, and infrequent verbs, where their subjunctive usage drops to 20.5%. Despite their overall variability with intensional subjunctive mood, the Stage 3 IntHSs were still 100% accurate in subjunctive production with *salir* and 87.5% accurate with *hacer*. In other words, their variability with subjunctive mood does not appear to have significantly affected their knowledge of subjunctive with certain, highly frequent Spanish verbs.

The final four IntHSs fell into Stage 4. Instead of categorically or variably producing subjunctive mood with *para que*, Stage 4 HSs developed an innovative grammar in which only indicative mood morphology follows the complementizer *para que*, regardless of the frequency of the verb forms that they are inflecting.

Figures 26 and 27 summarize the proportion of subjunctive production by the HSs in each of the four stages with both (a) frequent and (b) infrequent verbs. Notice that for both proficiency groups, HSs in Stage 2 and Stage 3 produce much more subjunctive with frequent verbs (black columns) than infrequent verbs (striped columns).

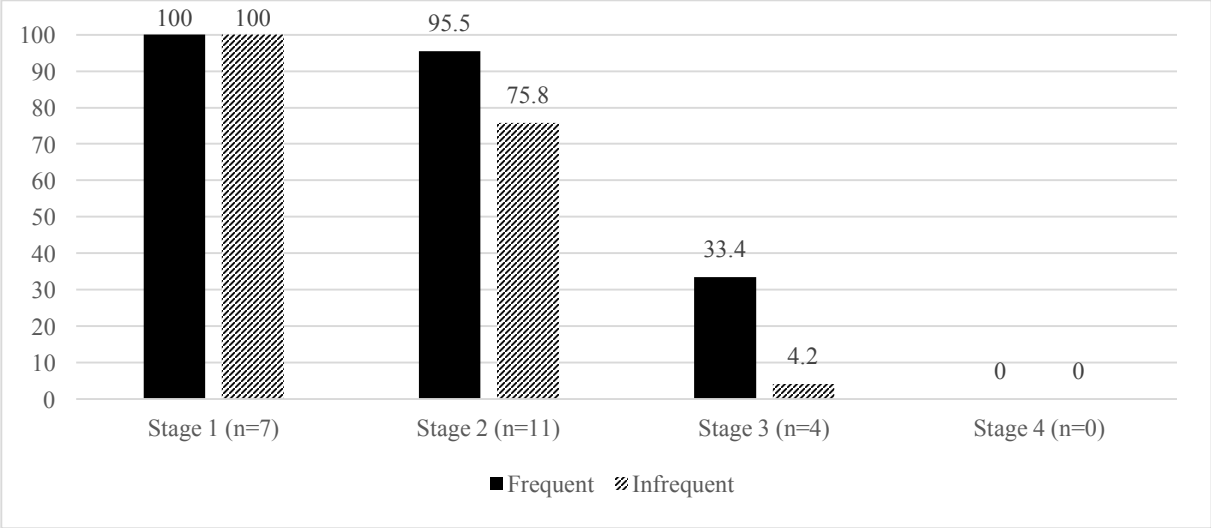

Figure 26: AdvHSs' use of intensional subjunctive: Stage x Frequency

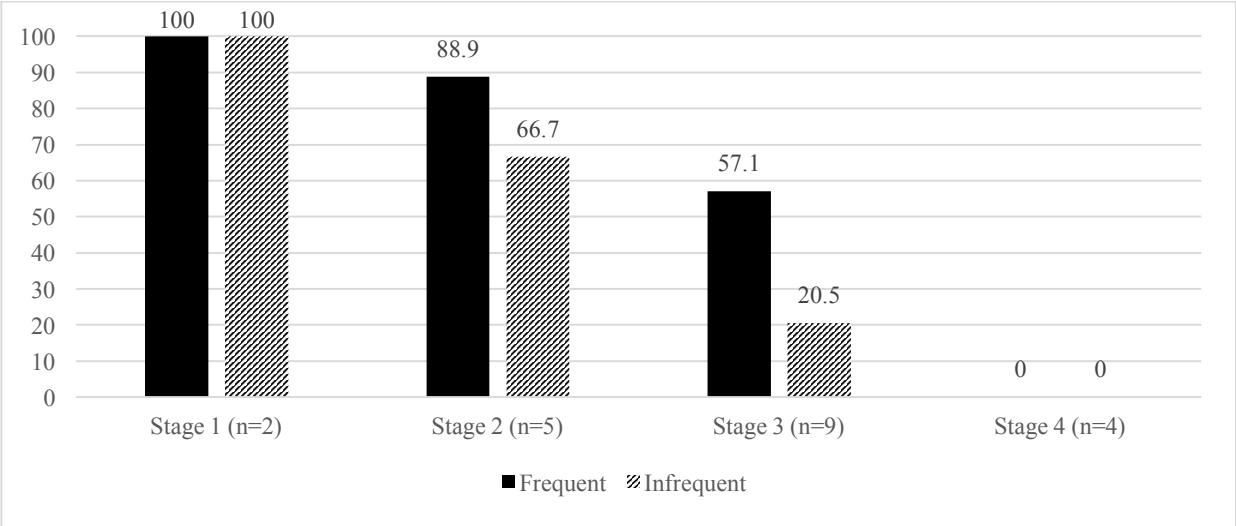

Figure 27: IntHSs' use of intensional subjunctive: Stage x Frequency

Is it possible, given the immense variability exhibited by the HSs in Stage 2 and Stage 3, that these speakers simply lack systematic knowledge of intensional subjunctive mood? In other words, is it the case that Stage 2 and Stage 3 HSs, rather than having a *lexically-limited* knowledge of subjunctive mood, simply produce subjunctive and indicative mood forms randomly? To rule out this possibility, I will now present a stage-by-stage analysis of the intensional subjunctive items in the Mood Preference Task.

As shown in Figures 28 and 29, the Stage 1 AdvHSs (97.1% accuracy) and Stage 1 IntHSs (100% accuracy) demonstrate ceiling performance with intensional subjunctive items on the Mood Preference Task (MPT). This is not surprising given that these speakers all produced subjunctive categorically in the CEPT.

Interestingly, the Stage 2 AdvHSs (92.7%) and Stage 2 IntHSs (96%) also exhibit near-ceiling performance in the MPT. Their high accuracy here suggests that they have systematic, qualitatively target-like knowledge of intensional subjunctive mood, despite their emerging, and lexically-influenced, variability in the production of these forms.

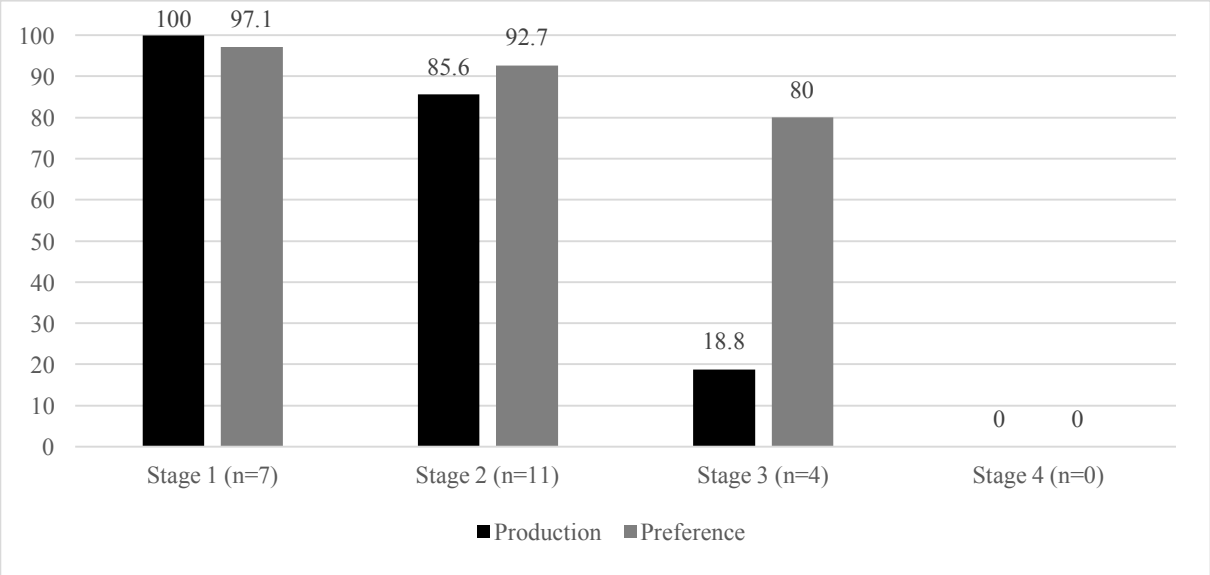

Figure 28: AdvHSs' accuracy with intensional subjunctive: Stage x Task

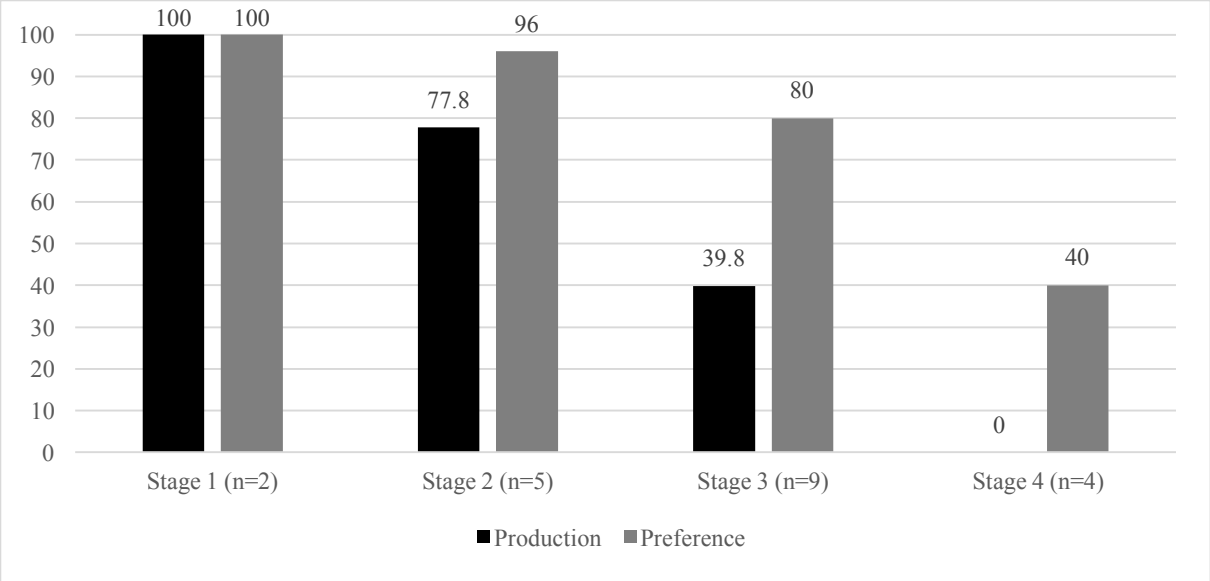

Figure 29: IntHSs' accuracy with intensional subjunctive: Stage x Task

More informative, perhaps, is the case of the Stage 3 HSs, who by definition produced intensional subjunctive mood less than half of the time. Recall that the Stage 3 HSs exhibited minimal production of subjunctive, almost all of which was limited to high frequency lexical items. Do these HSs, despite the lexically-limited nature of their subjunctive production, exhibit sensitivity to subjunctive mood in the MPT? Both the Stage 3 AdvHSs (80.0% accuracy) and the Stage 3 IntHSs (80.0% accuracy) perform well above-chance with intensional subjunctive in the MPT, indicating that beneath their productive variability lies systematic mood sensitivity. In this sense, the Stage 3 HSs closely fit Putnam & Sánchez's illustration of HSs in Stage 3: despite strongly reduced *production* of intensional subjunctive, they still seem to recognize it accurately, at least most of the time.

The Stage 4 HSs, all of whom were IntHSs, appear to be the only sub-group without systematic knowledge of intensional subjunctive mood. On the MPT, these speakers were 40% accurate, meaning that 60% of the time they preferred indicative mood after *para que*. It is possible, therefore, that these HSs' intensional mood knowledge is somewhat random.

Figure 30 shows the distribution of the AdvHSs with polarity subjunctive mood.

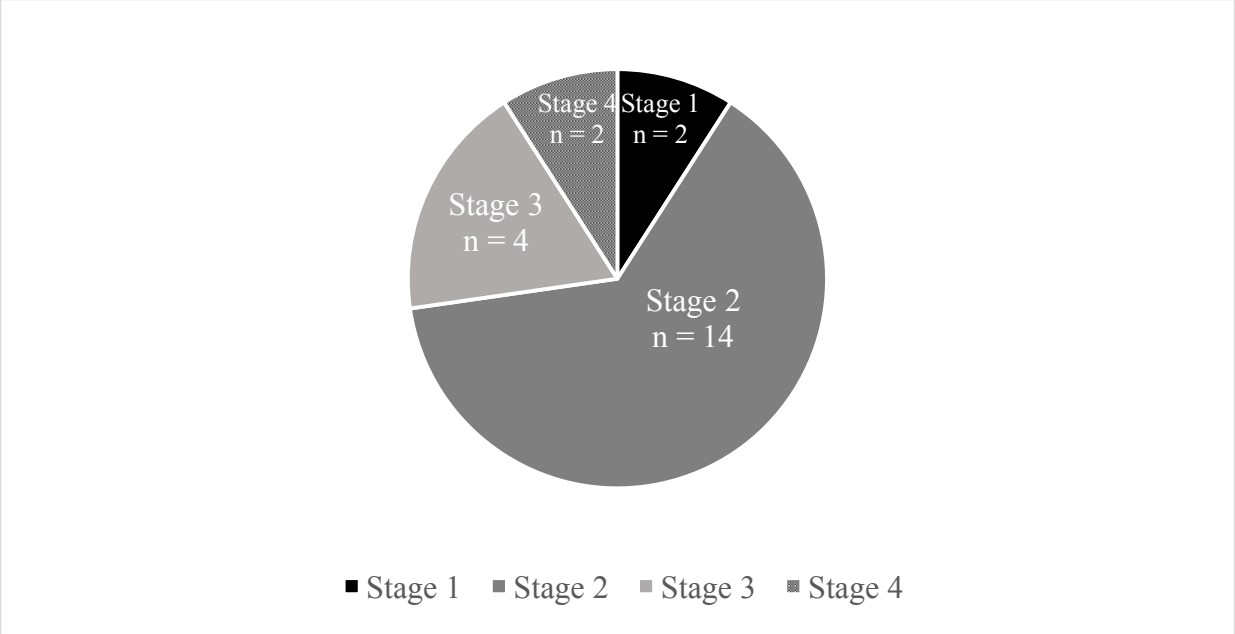

Figure 30: AdvHSs divided by stages: polarity subjunctive

Of the 22 AdvHSs, 16 (72.7%) fell into Stage 1 (n=2) or Stage 2 (n=14). As with intensional subjunctive, the Stage 2 AdvHSs showed nearly categorical production (90.4%) of polarity subjunctive with frequent verbs, including 100% subjunctive production with *tener*, 92.9% production with *decir* and *hablar* and 92.3% production with *llamar*. The Stage 2 AdvHSs, not surprisingly, exhibited more variability with infrequent verbs, which they inflected with subjunctive 77.8% of the time. Speakers in Stage 2 maintain strong knowledge of subjunctive but with a slightly reduced subset of Spanish verbs.

The remaining 6 AdvHSs (27.3%) fell into Stage 3 (n=4) or Stage 4 (n=2). Stage 3, once again, is defined by (a) a precipitous decline in overall subjunctive production relative to the previous Stages, as well as (b) a strong asymmetry between the production of subjunctive with frequent and infrequent verbs. Not surprisingly then, the four Stage 3 AdvHSs produced a much higher proportion of polarity subjunctive with frequent verbs

(45.8%) than infrequent verbs (20.8%). Interestingly though, these speakers produced polarity subjunctive only 25% of the time with the most frequent verb, *tener*, and only 50% with the second most frequent verb, *decir*. The final two AdvHSs were categorized as Stage 4 speakers since they did not produce any instances of polarity subjunctive.

Figure 31 shows the distribution of the IntHSs with polarity subjunctive mood.

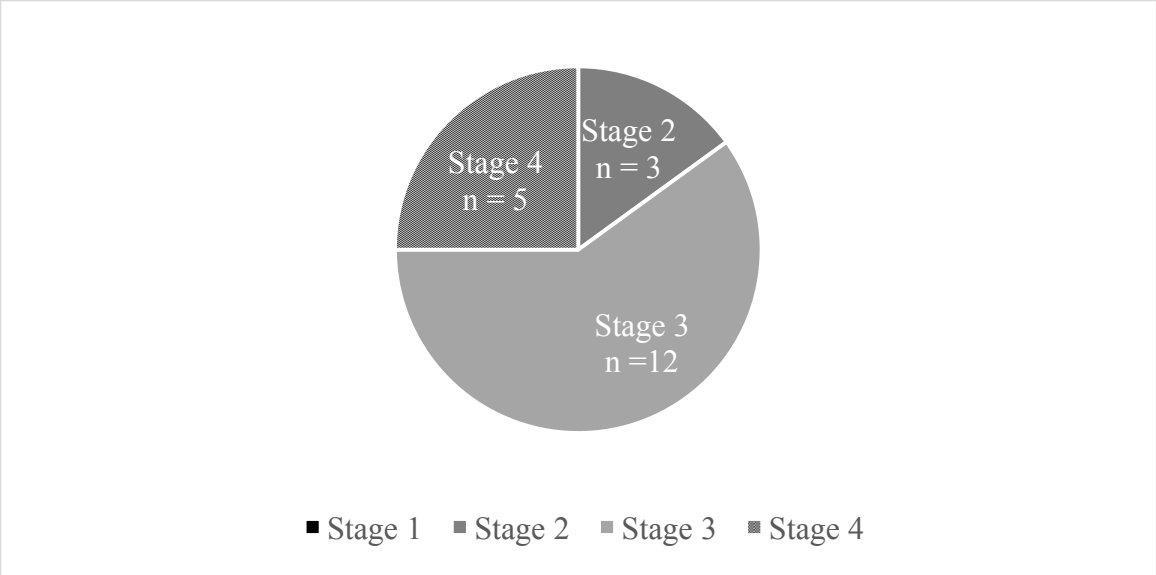

Figure 31: IntHSs divided by stages: polarity subjunctive

Only 3 of the 20 IntHSs (15.0%) fell into Stage 1 (n=0) or Stage 2 (n=3). The Stage 2 IntHSs produced polarity subjunctive 88.2% of the time with frequent verbs and 72.2% of the time with infrequent verbs, displaying the now-familiar pattern of mood variability with infrequent verbs. Tellingly, these speakers were 100% accurate in the production of the three most frequent verbs in the polarity subjunctive: *tener*, *decir* and *hablar*.

The vast majority of the IntHSs (85.0%) fell into Stage 3 (n=12) or Stage 4 (n=5) with polarity subjunctive. The Stage 3 IntHSs, like the other Stage 3 HSs examined thus far, exhibited much lower production of subjunctive mood than the speakers in Stage 1 or

Stage 2. That said, these speakers no longer showed a strong lexical frequency effect, producing subjunctive 33.3% of the time with frequent verbs compared to 29.9% of the time with infrequent verbs. Nonetheless, it is worth noting that the Stage 3 IntHSs were most accurate (63.6%) in the production of *decir*, which was the second most frequent verb. The five Stage 4 IntHSs did not produce any polarity subjunctive forms at all.

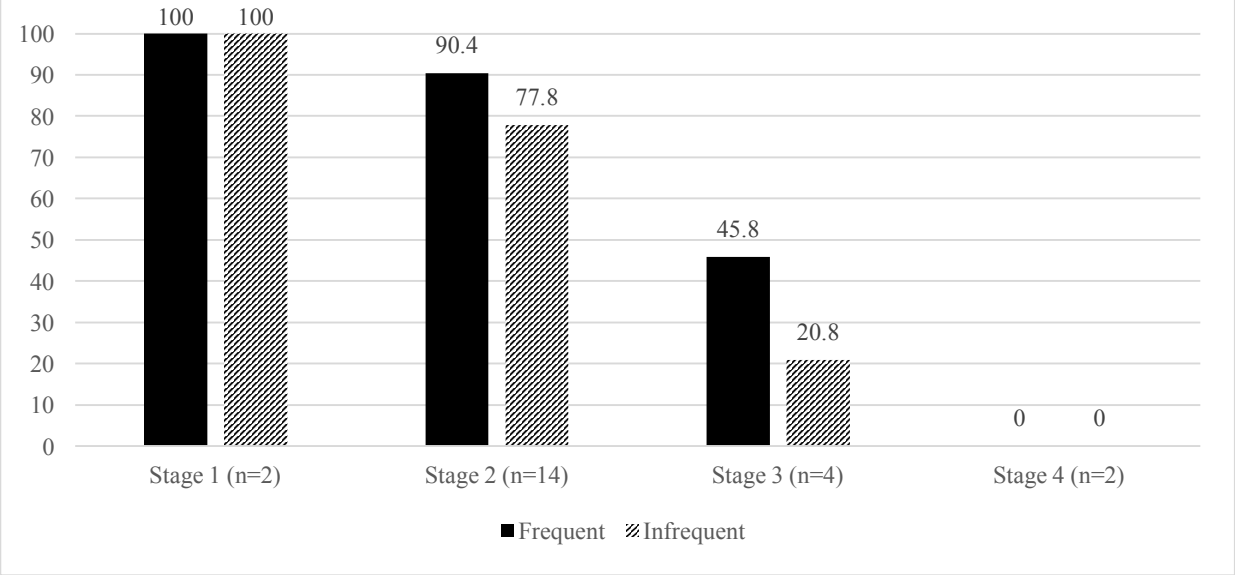

Figure 32: AdvHSs' use of polarity subjunctive: Stage x Frequency

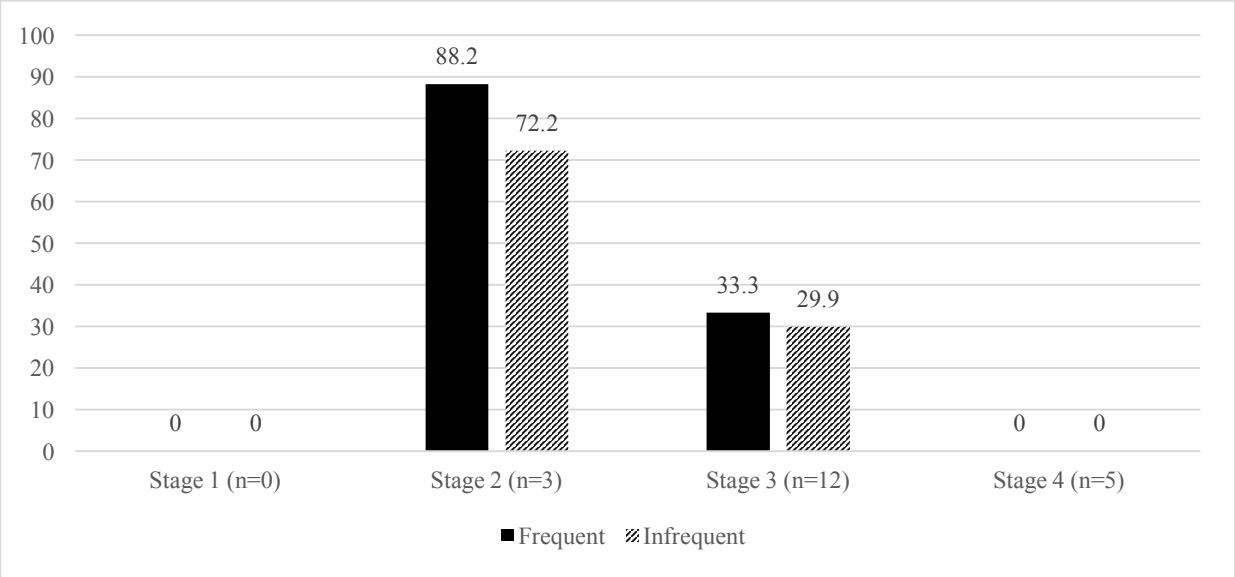

Figure 33: IntHSs' use of polarity subjunctive: Stage x Frequency

Figures 32 and 33 above summarize the proportion of subjunctive production by the HSs in each of the four stages with both (a) frequent and (b) infrequent verbs. Note that for both proficiency groups, HSs in Stage 2 and Stage 3 produce more subjunctive with frequent verbs (black columns) than infrequent verbs (striped columns).

It is possible, given HSs' high variability with polarity subjunctive production, that the HSs in Stage 2 and Stage 3 exhibit variability because they lack underlyingly systematic knowledge of polarity subjunctive mood in Spanish. To evaluate this possibility, we will now complement the productive analysis by considering the results of the Contextualized Acceptability Task (CAT) and the Mood Preference Task (MPT), both of which gauged HSs' *receptive* knowledge of polarity subjunctive mood.

Figure 34 shows the performance of the AdvHSs, divided into stages, on the CAT. The AdvHSs in all four stages rate (a) indicative items higher than subjunctive items in the Presupposition condition and (b) subjunctive items higher than indicative items in the No Presupposition condition, thereby indicating mood sensitivity. Notably though, the Stage 1 and Stage 2 AdvHSs (black and dark grey columns) make stronger mood distinctions in the No Presupposition condition than the Stage 3 and 4 AdvHSs. This finding suggests that grouping HSs into different stages of knowledge predicts their comprehension of mood.

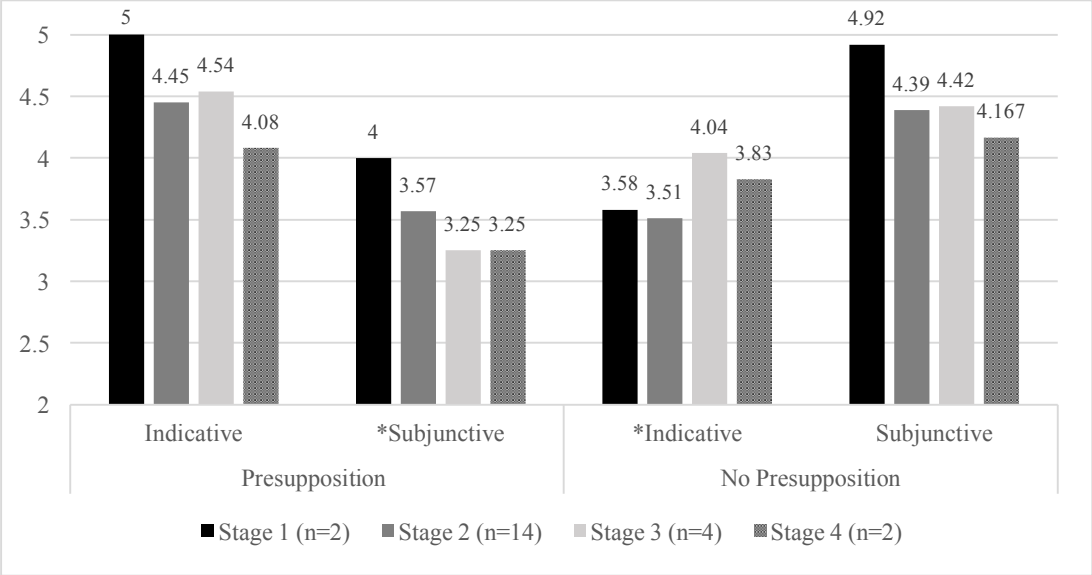

Figure 34: AdvHSs' acceptability ratings: Stage x Condition

Figure 35 shows the performance of the IntHSs, divided into stages, on the CAT. Not surprisingly, Stage 2, Stage 3, and Stage 4 IntHSs rate indicative mood higher than subjunctive mood in the Presupposition condition, as expected. Where these three stages of HSs differ, however, is in the No Presupposition condition, where subjunctive mood is the felicitous mood morphology. In this condition, the three Stage 2 IntHSs rate subjunctive items much higher than indicative items, showing target-like mood sensitivity. Stage 3 and Stage 4 IntHSs, on the other hand, demonstrate two types of divergent mood sensitivity. Stage 3 IntHSs rate subjunctive and indicative mood items equally while the Stage 4 IntHSs show a tendency to rate indicative items higher than felicitous subjunctive items. As in the case of the AdvHSs, the four HS stages are predictive of IntHSs' receptive mood sensitivity.

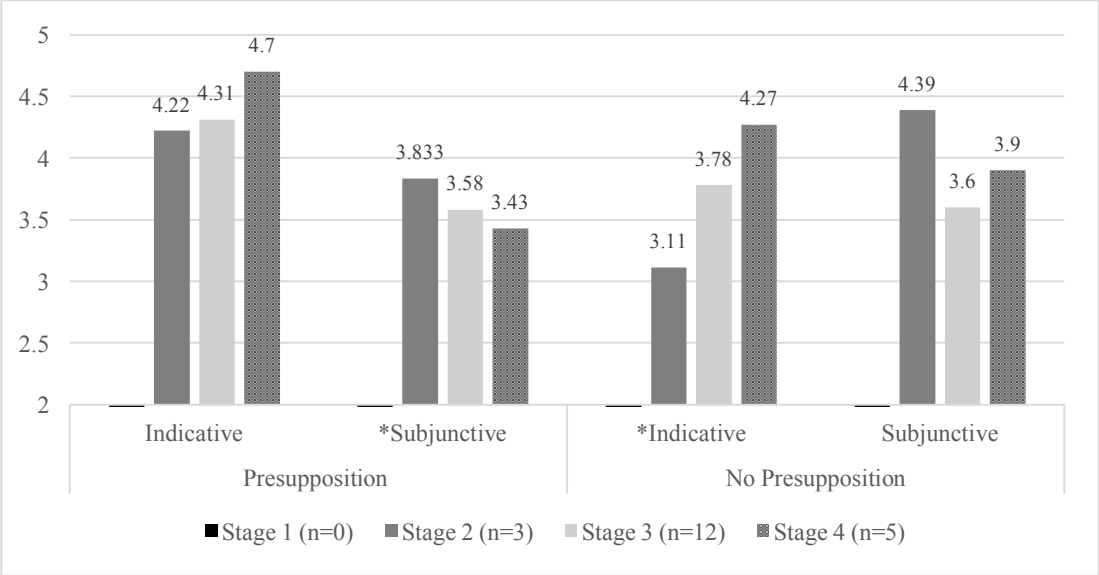

Figure 35: IntHSs' acceptability ratings: Stage x Condition

The results of the MPT, like the results of the CAT, indicate that HSs in Stages 1 and 2 have a more "target-like" knowledge of polarity subjunctive mood in Spanish.

Figure 36 (below) reveals that the Stage 1 and Stage 2 HSs perform well above chance with polarity subjunctive mood items in the MPT. Stage 3 HSs, however, appear to have reached a point at which both indicative and subjunctive mood morphology are equally preferable in the No Presupposition condition. (In fact, both the Stage 3 AdvHSs and the Stage 3 IntHSs were exactly 50% accurate in their mood preferences in this particular condition.) By Stage 4, both the AdvHSs (40% accuracy; 60% indicative preference) and the IntHSs (12% accuracy; 88% indicative preference) appear to prefer indicative mood as a near default form in non-presuppositional adjectival relative clauses.

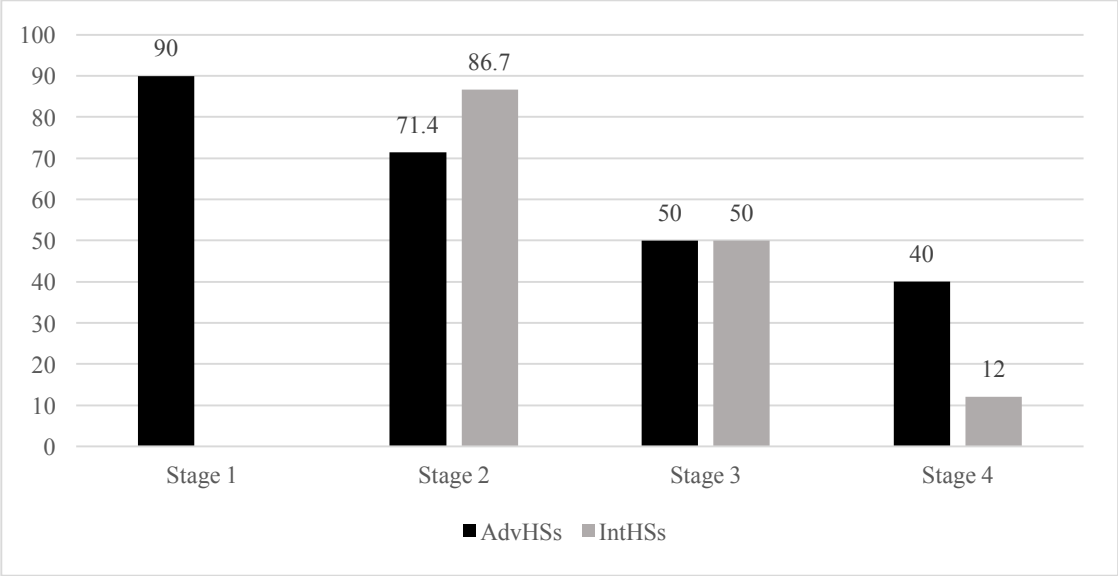

Figure 36: HSs' preference accuracy with polarity subjunctive: Stage x Proficiency

In summary, both group and individual results suggest that HSs' productive and receptive knowledge of subjunctive mood in Spanish is strongly modulated by lexical frequency. Though some HSs perform categorically in the production or comprehension of subjunctive mood, most exhibit differing degrees of variability which is concentrated in, though not exclusive to, their performance with lower frequency Spanish verbs.

6.3.3. Structural priming

Unlike the first two within-group factors mood selection type and lexical frequency, structural priming did not have any discernible impact on HSs' variability with subjunctive mood. What factors can explain the total lack of priming effects in the HSs? In this section, I will first focus on potential methodological problems before later evaluating how lexical frequency, as considered in Section 6.3.2, might also contribute to this finding.

In the polarity subjunctive condition, both the baseline and prime items required participants to read (out loud) a subjunctive mood form prior to spoken production. The

difference between the conditions, therefore, had to do with whether this previous subjunctive exposure was a *structural prime* or a *surface prime*. In the baseline condition, participants saw a surface prime, meaning that they were exposed to an intensional subjunctive structure (*querer* + *que* + subj.) before having to produce a polarity subjunctive form. Because this structure of subjunctive is hypothesized to involve different underlying syntactic features, it was not expected to prime participants. In the prime condition, on the other hand, participants saw a structural prime, meaning that they were exposed to an identical subjunctive structure before having to produce a polarity subjunctive form.

If mere exposure to *any* subjunctive mood morphology (e.g., regardless of the structural similarity) primed participants to produce subjunctive mood, then this could possibly account for why participants did not behave differently across the baseline and prime conditions since the statistical comparison would essentially be between two different prime conditions, rather than a prime condition and a true baseline. This possibility, however, does not account for why participants also showed no priming effect with intensional subjunctive, where the baseline conditions were "true baselines," e.g., not exposing participants to any subjunctive mood morphology. If exposure to *any* subjunctive mood boosted subjunctive production, then we would have expected to see a "priming" effect with intensional mood selection. There was, however, no such effect in the data.

The total lack of a priming effect creates a seemingly contradictory combination of experimental outcomes. First, HSs exhibited variability in the baseline conditions, which meant that they had room to "improve" their production of subjunctive. Second, HSs exhibited underlying knowledge of subjunctive (e.g., above chance performance), which suggested that they had the prerequisite linguistic representations needed to experience

priming ("improvement") upon exposure to that knowledge. (Priming effects are, after all, a test for the presence of underlying knowledge.) Third, despite these first two reasons, HSs did not produce any more subjunctive in the priming conditions.

How could such an unexpected outcome take place? One strong possibility is that priming effects were not observed due to differences in participants' lexical knowledge (Section 6.3.2), which could affect both their initial recognition of the primes as well as their production of the target verb forms. To illustrate this, I will now show what would have to happen for a participant to be primed in the CEPT. For the sake of simplicity, I will use an example from the intensional subjunctive condition, though the same logic would apply to the polarity subjunctive condition as well.

Structural Priming Example
ITEM #12 (Version A)
Context: Tu amiga no soporta el frío. Necesita una bufanda enorme para levantarse del sofá y salir de la casa. Le dices al dependiente:

"Busco una bufanda para que mi amiga se levante$_{SUBJ}$ del sofá y (SALIR)…"
(PRIME)

After listening to and reading the context, the participant begins to finish the incomplete sentence orally. When she arrives to the prime structure *para que se levante*, she must first recognize that the verb, *levante*, is inflected for subjunctive mood. It is conceivable, at this stage, that she might have underlying knowledge of subjunctive mood (e.g., as evidenced by an above chance production of subjunctive with *para que*) and yet not recognize this particular verbal form as pertaining to the class of subjunctive inflections, perhaps due to its relative infrequency. Consequently, she may not activate the underlying subjunctive features that she has, leading her to produce an indicative inflection

of the verb *salir*. In this case, the apparent lack of a priming effect is caused not by a deficit of featural knowledge, per se, but by the (perhaps temporary) inability to recognize that the prime verb is marked with subjunctive mood features.

It is important to note, at this point, how this first possibility distinguishes the priming element of this study, which focuses on morphological production, from that of priming studies which examine production of certain word orders (e.g., Loebell & Bock, 2003). In order to be primed by a given word order, a speaker only needs to recognize the lexical/semantic class of the priming elements. For example, to be primed into producing "double-object" constructions (e.g., '*Michael gave Stan the tuna'*) in English, speakers only need to recognize that *gave* is the main verb and *Stan* and *the tuna* are the two respective objects. In order to be primed by morphology, however, speakers must recognize the abstract features instantiated on the priming morpheme. If they do not, they will not show a priming effect, even if they "know" the abstract feature involved.

A second possible explanation for the non-priming effect has to do with HSs' knowledge of how subjunctive mood is instantiated on the target verb. It's logically possible that a HS could recognize the inflection on the prime verb *levante*, thereby triggering the activation of relevant subjunctive features in her mind, and yet not produce subjunctive on the target verb *salir* because she either does not know (or does not access) the subjunctive inflection of *salir* in real time. In this case, as in the first case, it would be a deficit in lexical, rather than morphosyntactic, knowledge which gives the (mistaken) impression that she does not have underlying knowledge of subjunctive mood.

Given these two logical possibilities, as well as the likelihood that HSs vary considerably in their ability to recognize and inflect subjunctive mood morphology

(demonstrated clearly in Section 6.3.2.), it is reasonable to conclude that the lack of a priming effect may be more of a lexical, rather than a morphosyntactic issue.

To conclude Section 6.3.3., we will now consider how the (non)priming effect ilustrated here might be explained by two of the three approaches to HS divergence/variability. Note that in the following paragraphs, I will exclude the Input Quality approach, as it does not make any identifiable predictions for priming.

The Representational Differences approach (Montrul, 2002, 2008, *inter alia*) could presumably attribute the lack of priming effects to representational differences in HSs' mood knowledge. That is to say, HSs may not respond to priming because they have an underlyingly different representation of the abstract features underlying mood and, consequently, cannot be primed by exposure to subjunctive forms. In light of the lack of priming effects, this account is reasonable (e.g., Jiang, 2000). Nonetheless, as we have seen in previous sections, arguing that HSs have different representations of mood leads to difficulties understanding their qualitatively target-like mood knowledge in multiple tasks.

The Activation/Lexicalist approach (Putnam & Sánchez, 2013) is best suited to explain the lack of priming effects observed in the present study, due to its emphasis on lexical knowledge and activation as a source of variability in HS grammars. As argued above, it is possible that HSs with underlying featural knowledge fail to show that knowledge in the present experiment due to difficulties either (a) recognizing the features present in the morphological primes or (b) knowing the morphological instantiation of mood on the target-verbs to be inflected in the task.

6.3.4. Summary of within-group factors

In Section 6.3, we have seen that both mood selection type and lexical frequency have a strong effect on HSs' variability in the production and comprehension of subjunctive mood. As with the predictive between-group factors, however, these two variables are not fully deterministic. It is not the case that HSs' subjunctive knowledge is limited to exclusively intensional, as opposed to polarity, selection or only frequent, rather than infrequent, verbs. It is best, therefore, to think of mood selection type and lexical frequency as factors which strongly shift the odds that HSs will exhibit subjunctive knowledge in a given productive or receptive context.

6.4. Modeling HSs' divergence and variability with subjunctive mood

In this last section of Chapter 6, I will conclude by achieving the following goals. In Section 6.4.1, I will use the results of the present study to model the shape of subjunctive mood development in the minds of Spanish HSs. In Section 6.4.2., I will elaborate on my four-stage model of HSs' mood variability, based on Putnam & Sánchez (2013).

6.4.1. Tracing subjunctive mood development

Between birth and age 5-6 (school age), HSs begin to learn Spanish, typically via naturalistic input received in the home. Their learning task, when it comes to subjunctive mood, is twofold. First, they must acquire the abstract features underlying mood. Previous research (e.g., Blake, 1983; Pérez-Leroux, 1998), considered alongside the findings of the present study, suggests that they will most likely acquire the *u*W feature of intensional subjunctive prior to the interpretable W feature of polarity subjunctive. Second, they must

learn to map these features, both in production and comprehension, to specific lexical items. Based on the findings of the present study, they will probably learn to do this first with more frequent verbs and then, maybe, infrequent verbs, too.

As HSs learn morphosyntactic features, as well as their mappings to specific lexical items, HSs with later AofA Eng will have more time for this subjunctive knowledge to "stabilize". HSs with earlier AofA Eng, on the other hand, will be more vulnerable to English influence, either directly (e.g., transfer of features from English onto HL lexical items) or indirectly (e.g., reduced relative usage of Spanish), leading to increased variability in their subjunctive production and comprehension. Among the HSs with earlier AofA Eng, it is likely that only those with frequent opportunities to use (as well as develop literacy in) Spanish will both (a) acquire the underlying mood features and (b) categorically map those features to lexical items in production and comprehension.

Between the ages of 7 and 17, as HSs progress through elementary, middle and high school, they will typically become more and more comfortable with English. This increased usage of English in school, as well as in social settings, will lead many HSs to exhibit increased variability with subjunctive mood over time, as documented by Merino (1983). For HSs with more opportunities to use Spanish at home and at school, this variability will likely be residual (e.g., the *occasional* production of indicative mood when subjunctive is expected), placing them firmly in the category of Stage 2 HSs. For HSs with fewer opportunities for Spanish use, however, this variability may take a more dramatic shape (e.g., occasional subjunctive production or, alternatively, indicative mood as default in all expected subjunctive contexts), marking them as clear examples of Stage 3 or 4 HSs. For all of these HSs, reductions in HL usage will almost certainly have a stronger impact

on their knowledge of polarity subjunctive, which is more complex to store and access for production and comprehension, as well as infrequent verbs, which are less frequently activated and less likely to be learned and accessed by HSs in real time.

Not all HSs, however, will exhibit significant subjunctive variability during this period. Some HSs, especially those with ample opportunities to use Spanish beyond the home, may come to *expand* their knowledge of subjunctive mood from the ages of 7 to 17. This expansion will be a function of (a) increased exposure to newer, less frequent subjunctive verb forms and (b) practice accessing these forms for production and comprehension. It is also possible that HSs with increased opportunities to use Spanish in the classroom will develop metalinguistic knowledge that further facilitates their production and comprehension of intensional and polarity subjunctive. As a result of their continued and varied opportunities to use Spanish, these HSs will likely reach advanced proficiency levels and, ultimately, perform like the Stage 1 HSs identified in Section 6.3.

Based on the results from the LCIs and SDCs, it is extremely unlikely that Spanish native speakers who immigrate to the US (a) between age 8 and age 12 or (b) at age 13 or later will exhibit subjunctive variability in expected subjunctive contexts, barring extreme circumstances such as adoption into an exclusively English-speaking family (e.g., Montrul, 2012). For these speakers, even sporadic use of Spanish seems to be more than enough to maintain stable intensional and polarity subjunctive knowledge.

At age 18 and beyond, it is likely that most HSs will continue to exhibit variability in their knowledge of subjunctive mood (Montrul, 2009; Montrul & Perpiñán, 2011; van Osch & Sleeman, *forthcoming*; Perez-Cortes, 2016). It is possible that this variability will fluctuate over time in accordance with HSs' changes in day to day Spanish usage. HSs who

move to monolingual English-speaking communities, for example, may begin to show more variability with subjunctive mood, possibly due to destabilization of their lexical knowledge. HSs who spend significant time abroad in Spanish-speaking countries or communities, however, may show significantly less variability over time. Given the dearth of longitudinal studies which investigate adult HSs' grammatical variability, it is unclear whether and to what extent HSs exhibit more or less grammatical variability over time.

There is very little about this account of HSs' mood development that would be disagreeable to proponents of the Input Quality and Representational Difference approaches, respectively. What sets the present account apart from these approaches is its focus on the HL lexicon as a primary locus of HS divergence and variability.

6.4.2. Four stages of HSs' subjunctive mood knowledge

To conclude Chapter 6, I will now briefly elaborate on my conceptualization of the different "stages" of HSs' subjunctive mood knowledge. As mentioned in Section 6.3.2., this idea of stages is based on Putnam & Sánchez (2013), though with a narrower overall scope. Instead of focusing on the *number* of FFs that transfer/reassemble from the dominant language to the HL, my own model addresses the *extent* to which a single FF in the HL exhibits variability, specifically as a consequence of differences in lexical frequency.

Figure 36 is a model of the four "stages" of HSs' subjunctive mood knowledge. For the sake of simplicity, this figure is not specific to either intensional or polarity subjunctive. The concentric circles represent the proportion of the HL verbal lexicon with which HSs can produce and comprehend subjunctive mood morphology. The larger circles represent higher proportions of the lexicon while smaller circles represent smaller proportions.

In Stage 1, HSs have underlying knowledge of the relevant FFs (e.g., *u*W or W) that underlie subjunctive mood. Crucially, this knowledge is not lexically limited, as these HSs know how the FFs map onto all verbal roots in Spanish. As a result, HSs in Stage 1 produce subjunctive mood categorically in expected contexts and understand it, too.

Like the Stage 1 HSs, Stage 2 HSs appear to have robust knowledge of the relevant FFs underlying subjunctive mood. However, Stage 2 HSs' featural knowledge is, in a broad sense, lexically limited. With a small but growing subset of HL verbal roots, Stage 2 HSs either (a) know how subjunctive mood is instantiated but do not always access its target form in real time or (b) do not know how subjunctive mood is instantiated morphologically. As a result, Stage 2 HSs produce subjunctive mood above chance, yet much more variably with less frequent verb forms. In comprehension, Stage 2 HSs will likely exhibit more target-like knowledge though here, as in the case of production, lexical knowledge could come to affect their ability to understand subjunctive mood on less frequent verb forms.

In Stage 3, HSs' subjunctive knowledge becomes significantly more difficult to define. On one hand, the fact that the Stage 3 HSs produce *some* subjunctive mood forms seems to suggest that they have underlying knowledge of the relevant FFs. That said, the vast majority of Stage 3 HSs' subjunctive mood production comes with very frequent verbs, pointing to the possibility that their subjunctive mood production is evidence of memorized chunks, rather than underlying featural knowledge. Unlike Stage 1 and Stage 2 HSs, Stage 3 HSs exhibit significant variability not just in production but also in *comprehension* of subjunctive mood, strengthening the possibility that their knowledge of the underlying FFs is not truly systematic. This possibility is represented in Figure 36 by the fact that the Stage 3 HSs' circle is not fully encapsulated by the circles of the Stage 1 and 2 HSs.

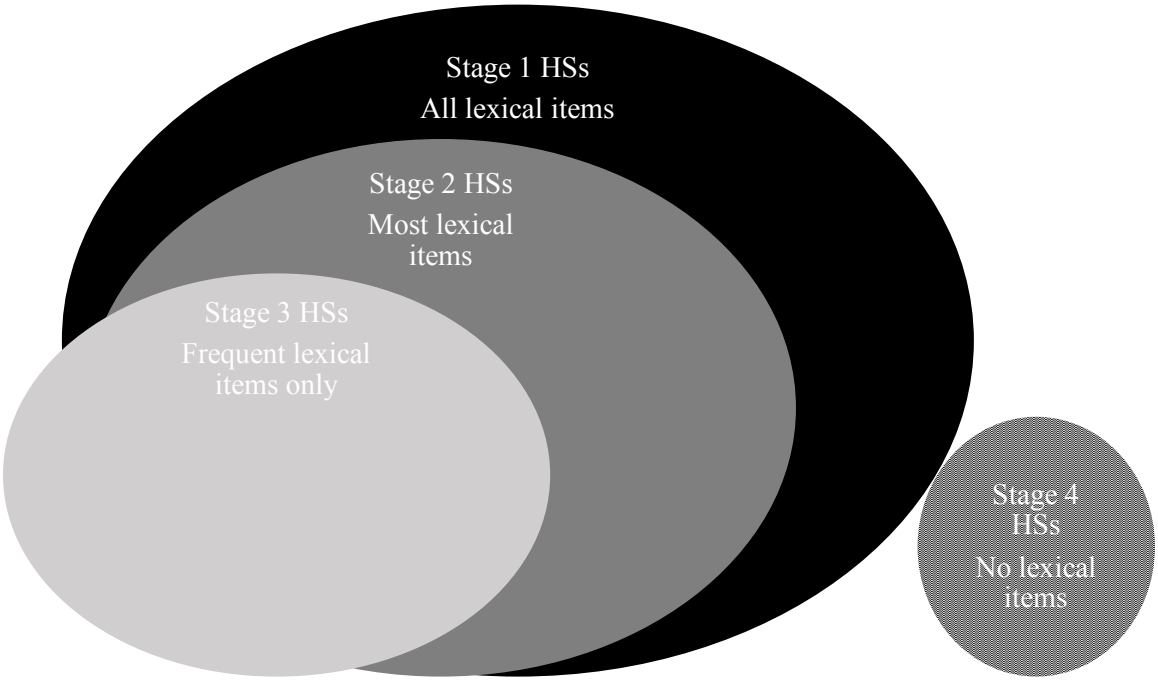

Figure 37: Conceptualization of HSs' stages of subjunctive mood knowledge

In Stage 4, the HSs no longer produce subjunctive mood morphology. Because they produce no subjunctive at all, they are represented in Figure 37 as existing outside the circles shared by HSs from the other three stages. Stage 4 HSs also do not comprehend subjunctive mood in receptive tasks, indicating that they (almost certainly) do not have knowledge of the FFs which underlie the use of subjunctive mood morphology in Spanish.

In Figure 36, I clearly place strong emphasis on the role of lexical knowledge/featural accessibility in HSs' subjunctive mood variability. Nonetheless, this conceptualization does not imply that *all* subjunctive mood variability is rooted in the HL lexicon. Stage 3 HSs, as presented in Section 6.3.2., exhibit high variability even with frequent verb forms, a finding which may be indicative of differences in these HSs' underlying featural or representational knowledge. (For a similar argument with gender

agreement in DPs, see Montrul & Potowski, 2007). The extent to which Stage 2 and Stage 3 HSs' subjunctive mood variability is triggered by (a) featural differences and (b) lexical differences remains an open, and relatively unexplored, question for future research.

CHAPTER 7:
CONCLUSION

7.1. Introduction

In this dissertation, I have examined HSs of Spanish and their production and comprehension of intensional and polarity subjunctive mood in Spanish. By exploring HSs' productive and receptive knowledge of intensional and polarity subjunctive mood, I have shed light on the nature of variability in HL acquisition. Specifically, I have identified a number of factors, both between-group and within-group, that strongly predict HSs' likelihood of exhibiting variable knowledge of subjunctive mood in Spanish. Of the four factors identified as significant predictors of HS variability with mood, the most notable is lexical frequency, which, to my knowledge, has not been systematically investigated in previous studies of HSs and their grammatical knowledge.

The remainder of this chapter will be organized as follows. In Section 7.2., I will summarize the main findings from the study, highlighting their relevance for HL theory, research methodology and pedagogy. In Section 7.3., I will discuss limitations of the study as well as some future directions for research on variability in HL acquisition.

7.2. Summary of main findings

In the present study, I found evidence of both HS divergence and HS variability (Chapter 2) with subjunctive mood morphology. By *HS divergence*, I mean to say that the HSs' knowledge of subjunctive differed from that of the SDCs. By *HS variability*, on the other hand, I mean to say that the HSs differed from *themselves*, e.g., alternately producing indicative and subjunctive mood morphology within given experimental conditions.

Though HS variability was more apparent in production, it was clearly observed in both productive and receptive experimental tasks.

The Contextualized Elicited Production Task (CEPT) tested HSs' oral production of subjunctive mood morphology. In the conditions of the CEPT where subjunctive mood was expected, many HSs exhibited variability by producing both subjunctive and indicative mood forms. The Contextualized Acceptability Task (CAT), on the other hand, tested participants' *comprehension* of mood morphology by assessing their ability to recognize (in)felicitous subjunctive and indicative mood forms in both [+presupposition] and [-presupposition] contexts, respectively. In the CAT, many HSs exhibited variability by accepting both felicitous *and* infelicitous mood forms within a given (non)presuppositional context. Lastly, the Mood Preference Task (MPT) tested participants' mood preferences by forcing them to select between minimal pairs of sentences which differed only in mood morphology. In the conditions of the MPT where participants were expected to prefer subjunctive mood, many HSs exhibited variability by preferring subjunctive mood with some experimental items and indicative mood with others.

Many previous studies have found evidence of HS variability with subjunctive mood morphology (e.g., Giancaspro, under revision; Montrul, 2007, 2009; Montrul & Perpiñán, 2011; van Osch & Sleeman, *forthcoming*; Pascual y Cabo, Rothman & Lingwall, 2012; Perez-Cortes, 2016). What sets the present study apart, however, is its fine-grained analysis of the different factors, both between-group and within-group, which predict HSs' likelihood of exhibiting such variability.

From a between-group standpoint, both (a) Spanish proficiency and (b) AofA Eng significantly affected HSs' subjunctive mood variability. Not surprisingly, advanced

proficiency HSs demonstrated less variability with subjunctive than intermediate proficiency HSs, a finding also reported by Giancaspro (under revision), Montrul (2009), Montrul & Perpiñán (2011), and Perez-Cortes (2016). With respect to AofA Eng, HSs with later AofA Eng exhibited less variability with subjunctive mood. While this AofA Eng effect has not previously been reported with mood morphology, AofA Eng has been shown to impact HSs' knowledge of other properties, including aspectual morphology in Spanish (Montrul, 2002) and binding properties in Korean (Lee, 2011).

From a within-group standpoint, two lexical factors emerged as predictors of HSs' subjunctive mood variability. The first factor was mood selection type. HSs exhibited significantly less variability with intensional subjunctive, which is lexically-selected, than polarity subjunctive, which is triggered by context. Following the analysis presented in Chapter 3, I argue that HSs' lower variability with intensional subjunctive is a result, at least partially, of differences in how intensional and polarity mood forms are stored in the mental lexicon. If HSs store the lexical selector *para que* ('so that') alongside its complementary structure (e.g., the *u*W feature in ForceP), this may make intensional subjunctive easier for HSs to access and produce. An additional explanation, which I cannot rule out here, is that intensional subjunctive with *para que* is earlier acquired.

The second within-group factor affecting HSs' subjunctive mood variability is lexical frequency, which played a significant role in both productive and receptive tasks. HSs were more likely to produce *and* understand subjunctive mood with frequent, as opposed to less frequent verbs. To my knowledge, this dissertation is the first experimental study to find that lexical frequency impacts the extent to which HSs exhibit morphological

variability. (For non-experimental observations about HSs and lexical frequency, see Dorian (1981) and Gal (1989)).

Of the three approaches to HS divergence and variability presented in Chapter 2, the Activation/Lexicalist approach (Putnam & Sánchez, 2013) is most consistent with the HSs' performance in this dissertation study. The key characteristic of this approach is its acknowledgement that functional features (FFs) may be acquired by a HS and yet, for different reasons (e.g., reduced HL activation), still not be categorically associated with or mapped to all lexical items in a given's HS's HL lexicon. This critical acknowledgement of the independence of FFs from lexical items (e.g., Adger, 2006; Adger & Smith, 2005, 2010; Nevins & Parrott, 2010, *inter alia*) provides this approach with the flexibility to account for non-binary outcomes in HL acquisition research, e.g., HSs who demonstrate systematic knowledge of subjunctive mood, but only with a limited subset of Spanish verbal forms. According to this approach, HSs exhibit such variability due to (a) difficulty accessing certain FFs on less frequent lexical items and/or (b) not having learned how such FFs are instantiated on certain, less commonly used lexical items in the first place.

It is not clear how the other two approaches to HS divergence and variability, previously presented in Chapter 2, might account for the findings of the present study.

The Input Quality approach (e.g., Pires & Rothman, 2009) suggests that HS variability, at least in some cases, is directly traceable to emerging variability in the input that HSs receive from first-generation immigrants. For reasons to be discussed in Section 7.3., I am unable to definitively rule out this possibility. Nonetheless, the (near) categorical subjunctive production of both the SDCs (and the LCIs) indicates that the HSs in the present study probably did not hear subjunctive variability from their parents, at least with

para que and in adjectival relative clauses. If the HSs did *not* hear subjunctive variability in the HL input, then their own mood variability must have another primary cause.

The Representational Differences approach (e.g., Montrul, 2002, 2008, *inter alia*) suggests that HSs' variability with a given property is evidence that their underlying grammatical representations of that property are qualitatively different from the representations of dominant speakers of the HL. This approach, without question, is appropriate for HSs who diverge sharply from HL-dominant speakers in both production *and* comprehension of a given property. It is almost certainly the case, for example, that five of the IntHSs in the study (specifically, those classified as "Stage 4" in Section 6.3.2.) do not have the same underlying representation of polarity mood as the SDCs. Not only did these five "Stage 4" speakers not *produce* any subjunctive mood, they also did not *understand* mood distinctions in both the CAT and the MPT, making it exceedingly unlikely that they possess the same representation of polarity mood as the SDCs.

In most instances, however, HSs in the present study diverged more subtly from the SDCs, producing subjunctive mood forms above chance *and* distinguishing, in the receptive tasks, between indicative and subjunctive mood morphology. These speakers' variable, yet systematic, performance with subjunctive mood morphology raises significant, and perhaps insurmountable, conceptual challenges for the Representational Differences approach. If these HSs are guided by a different or "incomplete," knowledge of subjunctive mood, as suggested by the Representational Differences approach, what makes them show the same grammatical tendencies as the SDCs, e.g., producing more subjunctive with *para que* than *porque* or rating subjunctive mood higher in the non-presupposition condition of the CAT? The Representational Differences approach, by

assuming that HS variability implies representational difference, fails to account for the results of these HSs whose grammatical knowledge is at once both variable *and* systematic.

Thus far, I have highlighted two achievements of the present dissertation. First, I have summarized factors which were found to predict HSs' subjunctive mood variability. Second, after an analysis of the variable patterns demonstrated by HSs, I have concluded that the results of the dissertation project are most consistent with the Activation/Lexicalist approach to HS divergence/variability. While each of these two achievements is valuable, the key contribution of this project is philosophical, rather than empirical in nature.

Up to this point, HL acquisition researchers have primarily focused on comparing HSs' knowledge of a given property to the knowledge of monolingual (e.g., van Osch & Sleeman, *forthcoming*) or bilingual (e.g., Perez-Cortes, 2016) control groups, respectively. The result of these comparisons, in most cases, is the same: HSs differ quantitatively, and sometimes qualitatively, from the controls in their knowledge of the property. In most studies, researchers follow up on this finding by showing that HSs with certain characteristics (higher HL proficiency: e.g., Montrul, 2009; later AofA Eng: e.g., Montrul, 2002; increased usage of the HL: e.g., Montrul & Sánchez-Walker, 2013; Perez-Cortes, 2016) differ less substantially from the control groups.

Without question, this line of inquiry is an important element of HL acquisition research, as it sheds light on the factors that make certain HSs more likely to differ from dominant speakers than others. That said, it is important to recognize that this common, comparative approach provides us with a less comprehensive view of HSs' grammatical knowledge. Comparing groups of HSs to groups of controls does not, and critically, *cannot*, tell us about the nature of the HS grammars themselves. Understanding HS grammars on

their own terms requires us to shift our focus from between-group differences to within-group differences (e.g., Bley-Vroman, 1983). Rather than comparing HSs to controls, we must compare HSs to *themselves*, investigating the linguistic factors that underlie their variable production and comprehension of different properties.

In the present dissertation, I take up this philosophical challenge by providing an in-depth focus on within-group differences. Like other studies of HSs' subjunctive knowledge (Perez-Cortes, 2016; van Osch & Sleeman, *forthcoming*), this dissertation explores the within-group variable of mood selection type, specifically by comparing HSs' likelihood of producing subjunctive in intensional and polarity mood conditions. Unlike previous studies, however, this project also explores the role of lexical frequency, which was found to significantly impact HSs' likelihood of producing subjunctive mood—*even within individual conditions* (e.g., intensional subjunctive). By illustrating the role of lexical frequency within a single experimental condition, the present study adds another previously unexplored dimension to our understanding of HSs' subjunctive mood systems.

The lexical frequency effect observed in the present dissertation also has implications for HL research methodology and design. If HSs' morphological knowledge is shaped, at least to some degree, by lexical frequency, then it is imperative that researchers be more deliberate about the lexical items that they choose to include in experiments with HSs. Had I ended up choosing only frequent verbs in my experimental tasks, I would have found significantly smaller differences between the subjunctive mood knowledge of the SDCs and the HSs. Alternatively, had I chosen only infrequent verbs, I would have found much larger differences. In either case, it's easy to see how failing to consider this variable could affect experimental results and the arguments we develop to explain them.

Researchers can easily address this problem by controlling for lexical frequency, especially when making within-group comparisons, such as the present study's comparison of intensional and polarity subjunctive mood selection. If there are substantial differences in lexical frequency across two within-group conditions, say, polarity and intensional mood selection, then the resulting confounded design will make it impossible to know whether any observed (lack of) effects are due to the variable of interest or lexical frequency. If lexical frequency is controlled, however, then the researchers will be able to state with confidence that the within-group effect observed is not attributable to lexical frequency.

Beyond contributing to HL acquisition theory and methodology, the present dissertation also has potential pedagogical implications for HL classrooms, which have rapidly increased in recent years (e.g., Carreira, 2013).

How, based on the results of the present study, should instructors teach subjunctive mood morphology in the HL classroom? One possible approach is to more carefully choose the *types* of subjunctive mood forms to which students are exposed, as well as the order in which they are exposed to them. Blake (1985), in an article about mood in the L2 classroom, argues that instructors should begin subjunctive mood lessons by focusing on categorical (e.g., non-variable) subjunctive uses, such as intensional subjunctive with *para que*, before later introducing more complex (and often variable) forms such as polarity subjunctive in adjectival relative clauses.

If such an approach were adopted in the introductory HL classroom, HSs would be able to start the often intimidating topic of subjunctive mood morphology by studying structures with which they are already (at least) somewhat familiar. In more advanced HL classes, where students possess much higher HL-proficiency, it may be more valuable for

teachers to take a different strategy, namely, dedicating the majority of their subjunctive mood instruction time to those polarity subjunctive structures known to be most variably produced by HSs. This reallocation of subjunctive class time may be most beneficial for HL classes with a strong presence of later childhood immigrants since, as shown in the present study, these bilinguals will have almost certainly maintained categorical knowledge of intensional subjunctive mood forms.

Given the finding that HSs' subjunctive production and comprehension is lowest with less frequent verbs, it may be similarly helpful for HL instructors to more thoughtfully consider the frequency of the verbs that they choose to give students for practice in and out of the classroom. An increased focus on practicing subjunctive mood with lower frequency verbs, for example, might help HSs to expand their knowledge of abstract grammatical categories, such as mood, to increasingly wider sets of HL lexical items.

7.3. Limitations and future directions

The present dissertation project has important implications for HL acquisition research. Nonetheless, there are limitations of the study which bear mentioning here.

Like most other research on adult HSs (for an exception, see Montrul & Bowles, 2010), the present study consists of a single experimental session, providing a snapshot of HSs' mood variability at a single point in time. Unfortunately, gauging HSs' grammatical knowledge in a single experimental session makes it more difficult to pin down the underlying causes of HS variability. Consequently, when a HS in this study does not produce subjunctive mood with a certain verb, I cannot know if this is due to (a) temporary

difficulty accessing functional features for production or (b) not knowing the way subjunctive mood is instantiated on that particular verb.

In order to distinguish between these possibilities, it would be useful to test the *same HSs* with the *same verbs* on multiple occasions. If a HS fails to produce subjunctive mood with a particular verb across multiple sessions, this would point towards lexical knowledge as the source of his mood variability with that verb. If, however, this HS produces both indicative *and* subjunctive forms of this verb across different sessions, this would point towards temporary production difficulties (or perhaps even representational differences) as the underlying source of his subjunctive mood variability with that verb. In any case, future research will certainly benefit from using experimental methods to test individual HSs repeatedly over time, ideally before and after these HSs have experienced significant increases in HL usage (e.g., study abroad). (See Chamorro, Sorace & Sturt, 2016 for a similar concept tested in a study of L1 attrition.)

A second limitation of the present study is the lack of online/processing data. Though oral production tasks, such as the CEPT, could certainly be considered online, none of the tasks in the present study directly evaluated HSs' processing of subjunctive mood. By not capturing participants' moment to moment production and comprehension, the present study is unable to provide a more thorough picture of the exact way that HSs activate and access subjunctive mood forms in real time. If, in a timed task, HSs were shown to be slower to produce or recognize subjunctive mood with infrequent verbs, this would constitute even stronger support for the Activation/Lexicalist approach presented throughout this dissertation. In the absence of such data, however, the support for this approach remains necessarily more speculative in nature. Ideally, to provide a more

complete picture of HSs' variability, future research should include both on-line and off-line measures of HSs' subjunctive mood knowledge.

A third limitation of the present study is its inability to adequately identify, and therefore account for, the potential effect of HSs' exposure to different input quality *from other HSs*. Based on the results from the SDCs, it is quite unlikely (though not out of the question; recall the results from Bookhamer, 2013) that HSs receive qualitatively different subjunctive mood input from their first-generation parents. However, HSs also receive Spanish input from English-dominant HSs, whose qualitatively different input can and likely does shape their subjunctive mood acquisition and maintenance.

The vast majority of the HSs in the present study have one, or many, siblings, most of whom are likely, based on the results of this study, to exhibit variability with subjunctive mood. HSs may, for example, hear siblings produce indicative mood morphology after *para que*. Exposure to such forms could play a role in triggering the extensive HS variability observed in the present study. Without testing HSs' siblings and other input sources, however, the present study leaves this potentially important variable unexplored.

There are at least two possible routes which researchers might take to more thoroughly account for the role of input quality in HS variability. The first possibility is to begin testing not just individual adult HSs but also their most common family interlocutors (e.g., parents, siblings, cousins…etc…). Taking this step, though tedious, would allow researchers to determine more confidently whether a given HS's variable production of a property is driven, at least in part, by exposure to similarly variable input from HS peers.

The second, and perhaps more promising, possibility is for researchers to test the impact of input quality in an experimental setting, e.g., by comparing the subjunctive mood

production of HS groups exposed to pre-production subjunctive mood stimuli which are either (a) invariant or (b) highly variable. Such experiments would shed light on whether, and to what extent, input quality truly affects HSs' variable grammatical production. Until such methodological steps are taken, it will be difficult to gauge the way in which HSs' input quality conspires with other linguistic and non-linguistic factors to create variability.

BIBLIOGRAPHY

Adger, D. (2006). Combinatorial variability. *Journal of Linguistics,* 42, 503-530.

Adger, D. & Smith, J. (2005). Variation and the minimalist program. In L. Cornips & K. Corrigan, *Syntax and Variation: Reconciling the Biological and the Social* (pp. 149-178). Amsterdam: John Benjamins Publishing.

Adger, D. & Smith, J. (2006). Variation in agreement: A lexical feature-based approach. *Lingua*, 120, 1109-1134.

Alba, R., Logan, J., Lutz, A., & Stults, B. (2002). Only English by the third generation? Loss and preservation of the mother tongue among the grandchildren of contemporary immigrants. *Demographics,* 39(3), 467-484.

Albirini, A. (2014). Toward understanding the variability in the language proficiencies of Arabic heritage speakers. *International Journal of Bilingualism*, 18(6), 730-765.

Albirini, A. & Benmamoun, E. (2014). Aspects of second language transfer in the oral production of Egyptian and Palestinian heritage speakers. *International Journal of Bilingualism*, 18(3), 244-273.

Albirini, A. & Benmamoun, E. (2015). Factors affecting the retention of sentential negation in heritage Egyptian Arabic. *Bilingualism: Language and Cognition*, 18(3), 470-489.

Albirini, A., Benmamoun, E. & Saadah, E. (2011). Grammatical features of Egyptian and Palestinian Arabic heritage speakers' oral production. *Studies in Second Language Acquisition*, 33, 273-303.

Anderson, R. (1999). Impact of first language loss on grammar in a bilingual child. *Communication Disorders Quarterly,* 21(1), 4-16.

Anderson, R. (2001). Lexical morphology and verb use in child first language loss: A preliminary case study investigation. *International Journal of Bilingualism*, 5(4), 377-401.

Austin, J., Blume, M. & Sánchez, L. (2015). *Bilingualism in the Spanish-speaking World.* Cambridge University Press: Cambridge, UK.

Bayram, F., Rothman, J., Iverson, M., Kupisch, T., Miller, D., Puig-Mayenco, E. & Westergaard, M. (2017). Differences in use without deficiencies in competence: Passives in the Turkish and German of Turkish heritage speakers in Germany. *International Journal of Bilingual Education and Bilingualism,*

Benmamoun, E., Montrul, S. & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, 39(3/4), 129-181.

Berman, R. (2004). Between emergence and mastery: The long developmental route of language acquisition. In R. Berman (Ed.), *Language Development across Childhood and Adolescence* (pp. 9-34). Amsterdam: John Benjamins.

Berman, R. (2007). Developing linguistic knowledge and language use across adolescence. In Hoff, E. & Shatz, M. (Eds.), *Blackwell Handbook of Language Development* (pp. 347-367). Malden: Blackwell Publishing.

Birdsong, D. (1999). *Second language acquisition and the critical period hypothesis.* Mahwah: Lawrence Erlbaum Associates.

Birdsong, D. (2009). Age and the end-state for second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 401-424). Bingley: Emerald Group.

Blake, R. (1982). El uso del subjuntivo con cláusulas nominales: regla obligatoria o variable. In H.L. Morales & M. Vaquero (Eds.), *Actas del primer congreso internacional sobre el español de América. Río Piedras* (pp. 351-360). Puerto Rico: Universidad de Puerto Rico.

Blake, R. (1983). Mood selection among Spanish-speaking children, ages 4 to 12. *The Bilingual Review*, 10, 21-32.

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

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning,* 33(1), 1-17.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, 20, 3-49.

Bookhamer, K. (2013). *The variable grammar of the Spanish subjunctive in second-generation bilinguals in New York City* (unpublished doctoral dissertation). The City University of New York, New York.

Borgonovo, C., Bruhn de Garavito, J. & Prévost, P. (2015). Mood selection in relative clauses: Interfaces and variability. *Studies in Second Language Acquisition,* 37, 33-69.

Bosque, I. (2012). Mood: Indicative vs. subjunctive. In J. Hualde, A. Olarrea & E. O'Rourke (Eds..), *The Handbook of Hispanic Linguistics* (pp. 371-394). Malden, MA: Blackwell Publishing.

Boumans, L. (2006). The attributive possessive in Moroccan Arabic spoken by young bilinguals in the Netherlands and their peers in Morocco. *Bilingualism: Language and Cognition*, 9(3), 213-231.

Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition,* 33, 247-271.

Bruhn de Garavito, J. (2002). Verb raising in Spanish: A comparison of early and late bilinguals. In B. Skarabela, S. Fish & A. Do (Eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development (BUCLD)* (pp. 84-94). Somerville, MA: Cascadilla Press.

Bylund, E. (2009). Maturational constraints and first language attrition. *Language Learning*, 59(3), 687-715.

Carreira, M. (2011). The vitality of Spanish in the United States. *Heritage Language Journal,* 10(3), 396-413.

Chamorro, G., Sorace, A. & Sturt, P. (2016). What is the source of L1 attrition? The effect of recent L1 exposure on Spanish-speakers under L1 attrition. *Bilingualism: Language and Cognition*, 19(3), 520-532.

Chang, C. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, 19(4), 791-809.

Correa, M. (2011). Heritage language learners of Spanish: What role does metalinguistic knowledge play in their acquisition of the subjunctive? In L. Ortiz-López (Ed.),

Selected Proceedings of the 13th Hispanic Linguistics Symposium (pp. 128-138). Somerville, MA: Cascadilla Proceedings Project.

Culicover, P. (2013). *Grammar and Complexity: Language at the Intersection of Competence and Performance.* Oxford: Oxford University Press.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research,* 49(2), 222-251.

Cuza, A. (2010). On the L1 attrition of the Spanish present tense. *Hispania,* 93(2), 256-272.

Cuza, A. (2013). Crosslinguistic influence at the syntax proper: Interrogative subject-verb inversion in heritage Spanish. *International Journal of Bilingualism,* 17(1), 71-96.

Cuza, A. (2016). The status of interrogative subject-verb inversion in Spanish-English bilingual children. *Lingua*, 180, 124-138.

Cuza, A. & Frank, J. (2011). Transfer effects at the syntax-semantics interface: The case of double-*que* questions in heritage Spanish. *Heritage Language Journal,* 8(1), 66-89.

Cuza, A. & Frank, J. (2015). On the role of experience and age-related effects: Evidence from the Spanish CP. *Second Language Research,* 31, 3-28.

Cuza, A. & Miller, L. (2015). The protracted acquisition of past tense aspectual values in child heritage Spanish. In R. Klassen, J. Liceras & E. Valenzuela (Eds.), *Hispanic linguistics at the crossroad: Theoretical linguistics, language acquisition and language contact* (pp. 211-230). Amsterdam: John Benjamins.

Cuza, A. & Pérez-Tattam, R. (2016). Grammatical gender selection and phrasal word order in child heritage Spanish: a feature re-assembly approach. *Bilingualism: Language and Cognition*, 19(1), 50-68.

Cuza, A., Pérez-Leroux, A.T., & Sánchez, L. (2013). The role of semantic transfer in clitic-drop among simultaneous and sequential Chinese-Spanish bilinguals. *Studies in Second Language Acquisition,* 35(1), 93-125.

Davies, M. (2006). *A Frequency Dictionary of Spanish: Core Vocabulary for Learners.* New York, NY: Routledge.

Diessel, H. (2007). Frequency effects in language acquisition, language use and diachronic change. *New Ideas in Psychology,* 25, 108-127.

Dorian, N. (1981). *Language death: the life cycle of a Scottish Gaelic dialect*. University of Pennsylvania Press: Philadelphia, PA.

Duffield, N. & White, L. (1999). Assessing L2 knowledge of Spanish clitic placement: convergent methodologies. *Second Language Research,* 15, 133-160.

Durlak, J. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, 34(9), 917-928.

Ellis, A. & Morrison, C. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 24(2), 515-523.

Ellis, A. & Lambon Rolph, M. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 26(5), 1103-1123.

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition,* 24, 143-188.

Ellis, N. & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form and function. Introduction to the special issue. *Modern Language Journal*, 93(3), 329-335.

Fernández Pedraza, M. (2014). *Variación sociosintáctica del modo subjuntivo en el español hablado en la ciudad de Milwaukee: Un estudio entre hispanohablantes de Puerto Rico* (unpublished Master's thesis). University of Wisconsin-Milwaukee, Milwaukee.

Fishman, J.A. (1964). Language maintenance and language shift as a field of inquiry: A definition of the field and suggestions for its further development. *Linguistics*, 2(9), 32-70.

Flores, C.M.M. (2015). Understanding heritage language acquisition: Some contributions from the research on heritage speakers of European Portuguese. *Lingua*, 164, 251-265.

Gal, S. (1989). Lexical innovation and loss: The use and value of restricted Hungarian. In N. Dorian (ed.), *Investigating Obsolescence* (pp. 313-331). New York: Cambridge University Press.

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

Garcia, C. (2011). Distinguishing two "synonyms": A variationist analysis of *quizá* and *quizás* in six Spanish dialects. In Michnowicz, J. & Dodsworth, R. (Eds.) *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics*. Cascadilla Proceedings Project: Somerville, MA (pp. 103-112).

Giancaspro, D. (under revision). Predicting variability in heritage speakers' subjunctive mood production.

Gollan, T., Montoya, R., Cera, C., & Sandoval, T. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58, 787-814.

Grüter, T. & Paradis, J. (Eds.) (2014). *Input and Experience in Bilingual Development*. Amsterdam: John Benjamins.

Gudmestad, A. (2010). Moving beyond a sentence-level analysis in the study of variable mood use in Spanish. *Southwest Journal of Linguistics*, 29(1), 25-51.

Guitart, J. (1982). On the use of Spanish subjunctive among Spanish-English bilinguals. *Word,* 1-2, 59-67.

Halle, M. & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S.J. Keyser (Eds..), *The View from Building 20: Linguistics Essays in Honor of Sylvain Bromberger* (pp. 111-176). Cambridge, MA: MIT Press.

Henry, A. (2005). Idiolectal variation and syntactic theory. In L. Cornips & K. Corrigan (Eds..), *Syntax and Variation: Reconciling the Biological and the Social* (pp. 109-122). Amsterdam: John Benjamins Publishing.

Hislope, K. (2001). *Reading and writing in Spanish by English-educated Spanish-English bilinguals* (unpublished doctoral dissertation). Purdue University, West Lafayette.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26, 55-88.

Hopp, H. & Putnam, M. (2015). Syntactic restructuring in heritage grammars: Word order variation in Moundridge Schweitzer German. *Linguistic Approaches to Bilingualism*, 5(2), 180-214.

Hulsen, M., de Bot, K., & Weltens, B. (1998). In T. Ammerlaan, M. Hulsen, H. Strating, & K. Yagmur (Eds..), *Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages*.Waxmann: New York (pp. 153-168).

Ionin, T. & Zyzik, E. (2014). Judgment and interpretation tasks in second language research. *Annual Review of Applied Linguistics*, 34, 37-64.

Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research*, 1146, 2-22.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434-446.

Jia, G. & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescentslearning English in the United States. *Applied Psycholinguistics*, 24, 131-161.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21(1), 47-77.

Johnson, J.S. & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology,* 21, 60-99.

Kaufman, D. (1998). Tales of L1 attrition: Evidence from pre-puberty children. In T. Ammerlaan, M. Hulsen, H. Strating, & K. Yagmur (Eds.), *Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages*.Waxmann: New York (pp. 185-202).

Kempchinsky, P. (2009). What can the subjunctive disjoint reference effect tell us about the subjunctive? *Lingua*, 119, 1788-1810.

Kim, K. (2014). *Unveiling linguistic competence by facilitating performance* (unpublished doctoral dissertation). University of Hawai'i, Manoa.

Kim, J.Y. (2016). *The perception and production of prominence in Spanish by heritage speakers and L2 learners* (unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana.

Kim, K., O'Grady, W. & Deen, K.U. (2014). Poor performance on scrambled Korean OSV sentences by Korean heritage children: Performance, not competence. In C. Chu, C. Coughlin, B.L. Prego, U. Minai, & A. Tremblay (Eds.), *Selected Proceedings of the 5th Conference on Generative Approaches to Language Acquisition North America* (GALANA 2012) (pp. 51-59). Somerville, MA: Cascadilla Proceedings Project.

Kirova, A. (2016). *Lexical and morphological aspects of gender and their effect on the acquistion of gender agreement in second language learners* (unpublished doctoral dissertation). Rutgers University, New Brunswick.

Kowal, J. (2007). La elección del modo subjuntivo en las subordinadas nominales. *Lingüística Española Actual,* 29(1), 45-74.

Krug, M. (2000). *Emerging modals: A corpus-based study of grammaticalization*. Berlin: Mouton de Gruyter.

Kupisch, T., Akpinar, D., & Stohr, A. (2013). Gender assignment and gender agreement in adult bilinguals and second language learners of French. *Linguistic Approaches to Bilingualism*, 3(2), 150-179.

Kupisch, T. & Barton, D. (2013). Generic reference in adult German bilinguals: how bilinguals deal with variation. *Studia Linguistica*, 67(1), 1-27.

Kupisch, T., Lein, T., Barton, D., Schröder, D., Stangen, I., & Stoehr, A. (2014). Acquisition outcomes across domains in adult simultaneous bilinguals with French as weaker and stronger language. *Journal of French Language Studies,* 24, 347-376.

Kupisch, T. & Rothman, J. (forthcoming). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*.

Kupisch, T. & van de Weijer, J. (2016). The role of the childhood environment for language dominance: A study of adult simultaneous bilingual speakers of German and French. In C. Silva-Corvalán and J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 174-194). Cambridge: Cambridge University Press.

Lantolf, J. (1978). The variable constraints on mood in Puerto Rican-American Spanish. In M. Suñer (Ed.), *Contemporary Studies in Romance Linguistics* (pp. 193-217). Washington, DC: Georgetown University Press.

Lardiere, D. (2005). On morphological competence. In L. Dekydspotter, R. Sprouse & A. Liljestrand (Eds..), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference* (GASLA 2004) (pp. 178-192). Somerville, MA: Cascadilla Proceedings Project.

Lastra, Y. & Butragueño, P. M. (2012). Aproximación al uso del modo subjuntivo en el *Corpus Sociolingüístico de la Ciudad de México. Boletín de Filología,* 47(2), 101-131.

Leal-Méndez, T., Rothman, J. & Slabakova, R. (2015). Discourse-sensitive clitic-doubled dislocations in heritage Spanish. *Lingua*, 155, 85-97.

Lee, T. (2011). Grammatical knowledge of Korean heritage speakers: early vs. late bilinguals. *Linguistic Approaches to Bilingualism*, 1(2), 149-174.

Lee, T. (2013). Variation among heritage speakers: Sequential vs. simultaneous bilinguals. *Journal of the National Council of Less Commonly Taught Languages,* 14, 1-26.

Levelt, W. (1989). *Speaking*. The MIT Press: Cambridge.

Lipski, J. (1993). Creoloid phenomena in the Spanish of transitional bilinguals. In A. Roca & J. Lipski (Eds.), *Spanish in the United States: Linguistic Contact and Diversity* (pp. 155-181). New York: Mouton de Gruyter.

Lohndal, T. & Westergaard, M. (2016). Grammatical gender in American Norwegian heritage language: Stability or attrition? *Frontiers in Psychology*, 7(344), 1-15.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition,* 12, 251-285.

López-Ornat, S. (1994). *La adquisición de la lengua española*. Madrid: Siglo XXI.

Lynch, A. (1999). *The subjunctive in Miami Cuban Spanish: Bilingualism, contact, and language variability* (unpublished doctoral dissertation). University of Miami, Miami.

Martillo-Viner, K. (2017). Second generation NYC bilinguals' use of the Spanish subjunctive in obligatory contexts. *Spanish in Context,* 13(3), 343-370.

Martínez-Mira, M.I. (2006). *Mood simplification: Adverbial clauses in heritage Spanish* (unpublished doctoral dissertation). University of Illinois, Urbana-Champaign.

McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research,* 24(4), 459-486.

McCarthy, C. (2012). Modeling morphological variation and development. *Linguistic Approaches to Bilingualism,* 2(1), 25-53.

Merino, B. (1983). Language loss in bilingual Chicano children. *Journal of Developmental Psychology*, 4, 277-294.

Mikhaylova, A. (2012). Aspectual knowledge of high-proficiency L2 and heritage speakers of Russian. *Heritage Language Journal*, 9(1), 187-206.

Mikulski, A. (2006). *Native intuitions, foreign struggles? Knowledge of the subjunctive in volitional constructions among heritage and traditional FL learners of Spanish* (unpublished doctoral dissertation). University of Iowa, Iowa City.

Mikulski, A. (2010). Age of onset of bilingualism, language use, and the volitional subjunctive in heritage speakers of Spanish. *Heritage Language Journal,* 7(1), 28-46.

Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition*, 5(1), 39-68.

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

Montrul, S. (2004b). Subject and object expression in Spanish heritage speakers: a case of morphosyntactic convergence. *Bilingualism: Language and Cognition,* 7(2), 125-142.

Montrul, S. (2007). Interpreting mood distinctions in Spanish as a heritage language. In K. Potowski & R. Cameron (Eds..), *Spanish in Contact: Policy, Social and Linguistic Inquirires* (pp. 23-40). Amsterdam: John Benjamins.

Montrul, S. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins Publishing.

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

Montrul, S. (2013). How "native" are heritage speakers? *Heritage Language Journal*, 10(2), 15-39.

Montrul, S. (2014). First language retention and attrition in an adult Guatemalan adoptee. In M. Schmid & B. Köpke (Eds..), *First Language Attrition* (pp. 91-126). Amsterdam: John Benjamins Publishing.

Montrul, S. (2016). *The Acquisition of Heritage Languages.* Cambridge: Cambridge University Press.

Montrul, S. (2016). Dominance and proficiency in early and late bilingualism. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 15-35). Cambridge: Cambridge University Press.

Montrul, S., Bhatt, R. & Bhatia, A. (2012). Erosion of case and agreement in Hindi heritage speakers. *Linguistic Approaches to Bilingualism*, 2(2), 141-176.

Montrul, S., Bhatt, R., & Ghirju, R. (2015). Differential object marking in Spanish, Hindi and Romanian as heritage languages. *Language*, 91(3), 564-610.

Montrul, S. & Bowles, M. (2010). Is grammar instruction beneficial for heritage language learners? Dative case marking in Spanish. *Heritage Language Journal,* 7(1), 47-73.

Montrul, S., Davidson, J., de la Fuente, I., & Foote, R. (2014). Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism: Language and Cognition,* 17(1), 118-138.

Montrul, S., Dias, R. & Santos, H. (2011). Clitics and object expression in the L3 acquisition of Brazilian Portuguese: Structural similarity matters for transfer. *Second Language Research,* 27(1), 21-58.

Montrul, S. & Foote, R. (2012). Age of acquisition interactions in bilingual lexical access: A study of the weaker language of L2 learners and heritage speakers. *International Journal of Bilingualism,* 18(3), 274-303.

Montrul, S., Foote, R. & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: the effect of age and context of acquisition. *Language Learning*, 58(3), 503-553.

Montrul, S. & Ionin, T. (2010). Transfer effects in the interpretation of definite articles by Spanish heritage speakers. *Bilingualism: Language and Cognition,* 13(4), 449-473.

Montrul, S. & Perpiñán, S. (2011). Assessing differences and similarities between instructed heritage language learners and L2 learners in their knowledge of Spanish tense-aspect and mood (TAM) morphology. *Heritage Language Journal*, 8(1), 90-133.

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

Montrul, S. & Sánchez-Walker, N. (2013). Differential object marking in child and adult Spanish heritage speakers. *Language Acquisition,* 20(2), 109-132.

Montrul, S. & Slabakova, R. (2003). Competence similarities between native and near-native speakers: An investigation of the preterite/imperfect contrast in Spanish. *Studies in Second Language Acquisition,* 25, 351-398.

Müller, N. & Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition,* 4(1), 1-53.

Nevins, A. & Parrott, J. (2010). Variable rules meet Impoverishment theory: Patterns of agreement leveling in English varieties. *Lingua,* 120, 1135-1159.

Nicoladis, E. (2006). Cross-linguistic transfer in adjective-noun strings by preschool bilingual children. *Bilingualism: Language and Cognition*, 9(1), 15-32.

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

O'Grady, W., Kwak, H.-Y., Lee, O. & Lee, M. (2011). An emergentist perspective on heritage language acquisition. *Studies in Second Language Acquisition,* 33, 223-245.

Oh, J. & Fuligni, A. (2010). The role of heritage language development in the ethnic identity and family relationships of adolescents from immigrant backgrounds. *Social Development*, 19(1), 202-220.

van Osch, B. & Sleeman, P. (forthcoming). Spanish heritage speakers in the Netherlands: Linguistic patterns in the judgment and production of mood. *International Journal of Bilingualism*.

Otheguy, R. (2014). The linguistic competence of second-generation bilinguals: A critique of "incomplete acquisition." In C. Tortora, M. den Dikken, I. Montoya & T. O'Neill (Eds..), *Romance Linguistics 2013: Selected Papers from the 43rd Linguistic Symposium on Romance Languages (LSRL), New York, 17-19 April, 2013* (pp. 301-319). Philadelphia: John Benjamins.

Parada, M. (2013). Sibling variation and family language policy: The role of birth order in the Spanish proficiency and first names of second-generation Latinos. *Journal of Language, Identity and Education*, 12, 299-320.

Pascual y Cabo, D. (2013). *Agreement reflexes of emerging optionality in heritage speaker Spanish* (unpublished doctoral dissertation). University of Florida, Gainesville.

Pascual y Cabo, D. (2016). Syntactic reflexes of emerging optionality in Spanish as a heritage language: The case of dative-experiencer verbs. *Hispania*, 99(1), 34-50.

Pascual y Cabo, D. & Gómez Soler, I. (2015). Preposition stranding in Spanish as a heritage language. *Heritage Language Journal,* 12(2), 186-209.

Pascual y Cabo, D. & Rothman, J. (2012). The (Il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics*, 33(4), 450-455.

Pascual y Cabo, D., Rothman, J. & Lingwall, L. (2012). Applying the Interface Hypothesis to heritage speaker acquisition: Evidence from Spanish mood. In A. Biller, E. Chung & A. Kimball (Eds..), *BUCLD 36: Proceedings from the 36th annual Boston University Conference on Languae Development* (pp. 437-448). Somerville, MA: Cascadilla Proceedings Project.

Perez-Cortes, S. (2016). *Acquiring obligatory and variable mood selection: Spanish heritage speakers' and L2 learners' performance in desideratives and reported speech contexts* (unpublished doctoral dissertation). Rutgers University, New Brunswick.

Pérez-Leroux, A.T. (1993). *Empty categories and the acquisition of wh- movement* (unpublished doctoral dissertation). University of Massachusetts, Amherst.

Pérez-Leroux, A.T. (1998). The acquisition of mood selection in Spanish relative clauses. *Journal of Child Language*, 25(3), 585-604.

Pérez-Leroux, A.T., Cuza, A. & Thomas, D. (2011). From parental attitudes to input conditions. In Potowski, K. & Rothman, J. (Eds.), *Bilingual Youth: Spanish in English-speaking Societies* (pp. 149-176). Philadelphia: John Benjamins.

Pérez Saldanya, M. (2012). Morphological structure of verbal forms. In J. Hualde, A. Olarrea & E. O'Rourke, *The Handbook of Hispanic Linguistics* (pp. 227-246). Malden, MA: Blackwell Publishing.

Pickering, M. & Ferreira, V. (2008). Structural priming: a critical review. *Psychological Bulletin*, 134(3), 427-459.

Pinker, S. & Ullman, M. (2002). The past and future of the past tense. *TRENDS in Cognitive Sciences,* 6(11), 456-463.

Pires, A. & Rothman, J. (2009). Disentangling sources of incomplete acquisition: an explanation for competence divergence across heritage grammars. *International Journal of Bilingualism*, 13(2), 211-238.

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

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

Polinsky, M. (2011). Reanalysis in adult heritage language: new evidence in support of attrition. *Studies in Second Language Acquisition,* 33, 305-328.

Polinsky, M. (2016). Looking ahead. In, D. Pascual y Cabo (Ed.), *Advances in Spanish as a Heritage Language* (pp. 325-346). Philadelphia: John Benjamins.

Portes, A. & Rumbaut, R. (2001). *Legacies: The Story of the Immigrant Second Generation.* Los Angeles: University of California Press.

Potowski, K. (2011). Identity and heritage language learners: Moving beyond essentializations. In S. Beaudrie & M. Fairclough (Eds.), *Spanish as a Heritage Language in the United States: The State of the Field* (pp. 179-199). Washington: Georgetown University Press.

Potowski, K., Jegerski, J. & Morgan-Short, K. (2009). The effects of instruction on linguistic development in Spanish heritage language speakers. *Language Learning,* 59(3), 537-579.

de Prada, A. & Pascual y Cabo, D. (2011). Invariable *gusta* in the Spanish of heritage speakers in the US. In J. Herschensohn & D. Tanner (Eds..), *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011*) (pp. 110-120). Somerville, MA: Cascadilla Proceedings Project.

Putnam, M. & Arnbjornsdottir, B. (2015). Minimizing (Interface) Domains: The Loss of Long-Distance Binding in North American Icelandic. In Page, B.R. & Putnam, M. (Eds.) *Moribund Germanic Heritage Languages in North America: Theoretical Perspectives and Empirical Findings* (pp. 203-223). Boston: Bill.

Putnam, M. & Salmons, J. (2013). Losing their passive voice: syntactic neutralization in heritage German. *Linguistic Approaches to Bilingualism*, 3(2), 233-252.

Putnam, M. & Sánchez, L. (2013). What's so incomplete about incomplete acquisition? A prolegemenon to modeling heritage language grammars. *Linguistic Approaches to Bilingualism*, 3(4), 476-506.

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

Quer, J. (2006). Subjunctives. In M. Everaert & H. van Riemsdijk (Eds..), *The Blackwell Companion to Syntax: Volume IV* (pp. 660-684). Maldan, MA: Blackwell Publishing.

Ramírez, D. (2007). *Language attrition and language maintenance: The Case of Colombian immigrants in New York state* (unpublished doctoral dissertation). University at Albany, State University of New York, Albany.

Rivera-Mills, S. (2011). Spanish heritage language maintenance. In S. Beaudrie & M. Fairclough (Eds.), *Spanish as a Heritage Language in the United States: The State of the Field* (pp. 21-42). Washington: Georgetown University Press.

Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: Inflected infinitives in heritage Brazilian Portuguese. *International Journal of Bilingualism,* 11, 359-389.

Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism,* 13(2), 155-163.

Rothman, J. & Treffers-Daller, J. (2014). A prolegemenon to the construct of the native speaker: heritage speaker bilinguals are natives too! *Applied Linguistics*, 35(1), 93-98.

Rouchdy, A. (1989). Urban and non-urban Egyptian Nubian: Is there a reduction in language skill? In N. Dorian (ed), *Investigating Obselescence*. Cambridge University Press: New York (pp. 259-266).

Rumbaut, R. & Massey, D. (2013). Immigration and language diversity in the United States. *Daedalus: Journal of the American Academy of Arts and Sciences*, 142(3), 141-154.

Rumbaut, R., Massey, D., & Bean, F. (2006). Linguistic life expectancies: Immigrant language retention in southern California. *Population and Development Review*, 32(3), 447-460.

Sánchez, L. (2004). Functional convergence in the tense, evidentiality and aspectual systems of Quechua-Spanish bilinguals. *Bilingualism: Language and Cognition,* 7(2), 147-162.

Sánchez Muñoz, A. (2007). *Register and style variation in speakers of Spanish as a heritage and as a second language* (Unpublished doctoral dissertation). University of Southern California, Los Angeles.

Sánchez Muñoz, A. (2009). *Spanish as a heritage language in the United States: A study of speakers' register variation.* Saarbrücken: Verlag Dr. Müller.

Schecter, S. & Bayley, R. (2002). *Language as Cultural Practice: Mexicanos en el Norte.* Mahway: Lawrence Erlbaum Associates.

Schmid, M. (2012). The impact of age and exposure on forgetting and retention of the birth language in international adoptees: a perspective from Holocaust survivors. *Linguistic Approaches to Bilingualism,* 2(2), 177-208.

Scontras, G., Fuchs, Z. & Polinsky, M. (2015). Heritage language and linguistic theory. *Frontiers in Psychology*, 6, 1-20.

Sharwood Smith, M. & van Buren, P. (1991). First language attrition and the parameter setting model. In H. Seliger & R. Vago (Eds.), *First Language Attrition*. Cambridge University Press: New York. (pp. 17-30).

Sherkina-Lieber, M. (2015). Tense, aspect, and agreement in heritage Labrador Inuttitut: do receptive bilinguals understand functional morphology? *Linguistic Approaches toBilingualism*, 5(1), 30-61.

Shin Kim, M. (2013). The mental lexicon of low-proficiency Korean heritage learners. *Heritage Language Journal*, 10(1), 17-35.

Silva-Corvalán, C. (1991). Spanish language attrition in a contact situation with English. In H. Seliger & R. Vago (Eds..), *First Language Attrition* (pp. 151-171). Cambridge: Cambridge University Press.

Silva-Corvalán, C. (1994). *Language Contact and Change: Spanish in Los Angeles*. Oxford: Clarendon.

Silva-Corvalán, C. (1994). Spanish language attrition in a contact situation with English. In H. Seliger & R. Vago (Eds.), *First Language Attrition*. Cambridge University Press: New York. (pp. 151-171).

Silva-Corvalán, C. (2014). *Bilingual language acquisition: Spanish and English in the first six years*. Cambridge: Cambridge University Press.

Silva-Corvalán, C. & Treffers-Daller, J. (Eds.) (2016). *Language dominance in bilinguals: Issues of measurement and operationalization*. Cambridge: Cambridge University Press.

Sorace, A. & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22(3), 339-368.

Sturt, P., Keller, F., & Dubey, A. (2010). Syntactic priming in comprehension: parallelism effects with and without coordination. *Journal of Memory and Language,* 62, 333-351.

de Swart, P. (2013). A (single) case for heritage speakers? *Theoretical Linguistics*, 39(3-4), 251-258.

Taylor, P., López, M.H., Martínez, J. & Velasco, G. (2012, Apr. 4). IV. Language use among Latinos. *Pew Research Center.* Retrieved from: http://www.pewhispanic.org/2012/04/04/iv-language-use-among-latinos/

Toribio, J. (2000). Setting parametric limits on dialectal variation in Spanish. *Lingua,* 10, 315-341.

Torres, J. (2013). *Heritage and second language learners of Spanish: The roles of task complexity and inhibitory control* (unpublished doctoral dissertation). Georgetown University, Washington, DC.

Valdés, G. (2000). Introduction. In *Spanish for Native Speakers, Volume I. AATSP Professional Development Series Handbook for Teachers K-16* (pp. 1-20). New York, NY: Harcourt College.

Valenzuela, E., Iverson, M., Rothman, J., Pascual y Cabo, D., Borg, K. & Pinto, M. (2015). Eventive and stative passives and copula selection in Canadian and American heritage speaker Spanish. In I. Pérez-Jiménez, M. Leonetti & S. Gumiel-Molina (Eds..), *New Perspectives on the Study of Ser and Estar* (pp. 267-292). Amsterdam: John Benjamins.

Wulff, S., Ellis, N., Römer, U., Bardovi-Harlig, K. & Leblanc, C. (2009). The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *Modern Language Journal,* 93(3), 354-369.

Yip, V. & Matthews, S. (2000). Syntactic transfer in a Cantonese-English bilingual child. *Bilingualism: Language and Cognition,* 3(3) 193-208.

Yip, V. & Matthews, S. (2007). *The Bilingual Child: Early Development and Language Contact.* Cambridge: Cambridge University Press.

Zentella, A. (1997). *Growing up bilingual.* Malden, MA: Blackwell Publishing.

Zobl, H. & Liceras, J. (2006). Competing grammars and parametric shifts in second language acquisition and the history of English and Spanish. In D. Bamman, T.

Magnitskaia, & C. Zaller (eds.), *Proceedings of the 30th Boston University Conference on Language Development (BUCLD)* (pp. 713-724). Somerville, MA: Cascadilla Press.