De-centering the Monolingual: A Psychophysiological Study of Heritage Speaker Language Processing

Christen N. Madsen II

The Graduate Center, City University of New York
DE-CENTERING THE MONOLINGUAL:
A PSYCHOPHYSIOLOGICAL STUDY OF HERITAGE SPEAKER LANGUAGE PROCESSING

by

Christen N Madsen II

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2018
De-centering the monolingual:

A psychophysiological study of heritage speaker language processing

by

Christen N Madsen II

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date

Gita Martohardjono
Chair of Examining Committee

Date

Gita Martohardjono
Executive Officer

Supervisory Committee:

Gita Martohardjono
Martin Chodorow
Richard G. Schwartz
De-centering the monolingual: A psychophysiological study of heritage speaker language processing

by

Christen N Madsen II

Advisor: Gita Martohardjono

Models of grammar, processing and acquisition are primarily built on evidence from monolinguals and adult learners of a second language. Heritage speakers, who are bilinguals of a societal minority language, acquire and use their heritage language in informal settings; but who live, work, and are educated in the societal majority language. The differences between heritage speakers and both monolinguals and adult second language learners are extensive: heritage speakers are not educated in the heritage language, their input is typically not from a prestige variety of the heritage language, and they are dominant in the majority language, using it more frequently (Valdés, 1989). Previous research of heritage speaker characterized their grammars as simple, decayed/attrited, and incomplete (Benmamoun, Montrul, & Polinsky, 2010; Scontras, Fuchs, & Polinsky, 2015), and are compared to intermediate second language learner grammars (Montrul, 2005).

The present study: 1) explores the language use and exposure of heritage speakers, 2) examines their performance on metalinguistic tasks, and 3) measures language processing using implicit measures (event-related potentials and pupillometry). Heritage speakers are compared to adult late second language learners living and working in a second language dominant society from the same community. The study focuses on fluent Spanish and English Latinx bilinguals living in the anglophone US. Spanish heritage speakers are appropriately compared to their time-apparent parents (English speaking Latinx immigrants who moved to the anglophone US in adulthood). Online language processing of subject- and object-relative clauses are examined as the subject-object relative clause processing asymmetry has been well-established in both Spanish and English, is early acquired, and is not confounded by prescriptive rules or literacy.
This dissertation is dedicated to the women in my family who were denied access to education. To my grandmother, Bertha García Serrano, who was not allowed to go to school after the fourth grade because "women don’t need school because their husbands will take care of them". To my mother, Betty Serrano Madsen, and my aunt, Sylvia Livingston, who never stopped trying to get more education in spite of all the barriers.
ACKNOWLEDGEMENTS

The dissertation was an enormous amount of work and is the culmination of the 16+ years of college. It represents more than just the contributions I am able to make to science with this document, but the capstone of all the work and experiences it took to get here. For that reason, I want to take the space to express my appreciation to everyone to helped me get to this point.

It has been a long road from night classes at a community college to a good job and a PhD across the country. Homelessness, illness, being kicked out a Mormon university for being gay, stopping during my BA due to lack of money, working while in school, being a first-generation college student. Any one of these likely could have stopped me from getting a PhD. While this is a personal triumph, it was only possible because of the immense support, encouragement, and kindness of my friends and family.

Thank you.

Likewise, all the determination and smarts in the world mean nothing without someone to show you the way like the wonderful mentors and professors I had.

Thank you.

I am very grateful to my committee at the CUNY Graduate Center, Gita Martohardjono, Martin Chodorow, and Richard G Schwartz, who more than anything, empowered me. They allowed me to own my work and gave me space to grow my skills and expertise and push myself. Their support created a safe space so that I felt like I could try new things. Not only did they provide guidance and support in terms of linguistic research, their support was instrumental in preparing me for a career outside of the Academy.

Gita and Richard were more than just professors I took classes from or whose lab I was in. I was fortunate to be able to work with Gita and Richard as a collaborator, co-author, co-presenter, and data analyst on the Second Generation Bilinguals Project (SGBP) in the Gita’s lab. Over these past 4 years, I learned more as a sort of apprentice watching them closely. I was able to see their passion for the work
and the desire to have our research have an impact. I am also grateful for the data I used in my dissertation. They generously let me carve out my own project and let me say what I wanted to. I am appreciative that they let me make this project my own.

From day one, eight years ago, Gita has been the supportive of me and I couldn't hope for a better advisor. Our relationship started out as casual conversations about life, then she became my advisor when I discovered I was more interested in quantitative research than phonology. She hired me in her lab, and in addition to providing financial support, the job provided a space that I was able to cultivate my interests in statistics and research design. Through the work in the lab, I was exposed to heritage speaker research which I now feel super passionate about. She has been a wonderful advisor by encouraging me to think and work strategically, be more focused, and to think of the big picture. I learned so much from watching her plan and think. She was always honest in her appraisal which was appreciated and helpful. Often times it can feel like linguistic research has no practical applications, but Gita melds the two. It was fulfilling to work with someone who knows and cares that what she researches should have an impact and finds a way to disseminate, advocate, and implement it. Lastly, and perhaps most importantly, I am grateful that I had an advisor that reminded me that there are things other than research that are important too, such as food, travel, dating, and dancing.

From Richard I learned how to think like a scientist. I learned to do the work first and crucially be able to support the ideas and the directions you want to go. This helped me to more clearly articulate my rationale and motivations explicitly. It also fundamentally changed the way I interact with people even outside of the Academy by being clearer about my assumptions when drawing any conclusion. I, and I'm sure the people around me, are grateful for this.

Martin was a wonderful mentor and sounding board to discuss statistics, which was my favorite part of this work. I learned so much from him and am always in awe watching him think. He is math goals. His willingness to just talk about stats was my favorite and I loved every meeting I had with him. I always learned something from Martin, even if it's listening to the same lecture I've heard 3 times already as his TA. In addition to being an advisor, Martin was instrumental in setting me on my current career path. When I first took statistics with Martin 7 years ago, it was a game changer. I was terrified and studied very hard. After that class, Martin was the first person I had heard consider me to be good at statistics, when a
student was asking for a tutor. Hearing this gave me the confidence to start to develop the skills in statistics and R programming. I cannot express enough how pivotal that compliment was as it set me on a path I love.

In addition to my dissertation committee, two professors at California State University, Fullerton were key in helping on my academic/professional journey. Cheryl Boyd Zimmerman gave me thoughtful advice that has stayed with me and was very helpful. Additionally, I would not have been able to enroll in a doctoral program without the kindness she showed me. My first advisor and mentor, Patricia Schneider Zioga, was the first person to suggest to me that I could get a doctorate when I was an undergrad. She helped me navigate my entry into the academy though difficulties.

I started my long academic journey at Riverside Community College and Norco College. At times it has felt like there are very few community college students who go on to earn doctorates. I grateful to the faculty at these colleges who provided high quality education. I never once felt like they treated their students as anything but capable.

I am blessed that I was able to work alongside Jennifer Chard, my dissertation buddy, colleague and dear friend. She was my accidental mentee in my second year, and we’ve been hustling side-by-side ever since. It is magical when you click with someone who has similar goals and work ethic. I learned so much from her about how to work, and work smart. I am lucky to have a colleague that compliments me so well. Her strengths, like lit reviews, are my weakness. She helped me lean into my strengths like realizing I am better at talking things out first. For the past 4 ½ years, we’ve met weekly to plan our personal and professional goals, and talk things out. Everything in this dissertation was first discussed with her. She was the sounding board when I just had to talk the narrative out before writing.

I have been a member of the Second Language Acquisition Lab (SLAL) for the past 5 years. The SLAL not only gave me a space to work with people, but also to develop skills outside of the PhD. I indebted to Ian Phillips, who I worked alongside as a co-lab manager for 3 years, for developing and managing SGBP. I would not have any data to analyze had he not been so organized and diligent in driving the
project forward as well as preprocessing all the data. I am thankful to all the RAs who ran participants. The dissertation would have been impossible without their hard work. Thank you Lianye Juliet Zhu, Reid Vancelette, Matthew G. Stuck, João Pedro Marinotti, Armando Tapia, Pamela Franciotti, LeeAnn Stover Stevens, and Cass Lowry.

In the course of the PhD program, a number of people in Linguistics had a profound impact on me. I saw them as a kind of big sister/brother model whether they knew it or not. I learned from watching them work on their dissertation, defend, navigate teaching, and have success in academic and non-academic careers. Thank you Eefje Boef, Matt Barros, Carolina Barrera Tobón, Elizabeth Pratt, Emily Zane, Marisa Nagano, Michaël Gagnon, and Jelena Runic.

During the PhD program I began to do freelance statistics and research design consulting. I am grateful not only for the work, but also to have learned from being by their side. Working with these clients and seeing them through different stages of the dissertation first ultimately made this task doable. Seeing every step of finishing a PhD through someone else's eyes demystified the processes and allowed me to feel confident that I too could finish the PhD. Thank you Chaya Nove and Maayan Barkan. Thank you Ingrid T Heidrick who I met with weekly for 2 years. I learned how to craft a story by watching her work on her dissertation from proposal to deposit. Thank you Myrtle Dickson who impressed me with can-do, get-it-done drive. Lastly, a special thanks to my first client, Euna Cho. I will always be grateful that she let me help her. I had wanted to get better at stats by doing and however incredulous she was that someone would WANT to do statistics, she let me. She opened the door to path I'm on now. The first drink will always be on me.

I am happy to have a number of friends who are also working on or have finished doctorates. Working side-by-side felt like we were in this together. I felt supported by these friends who knew exactly what it was like. Working alone at my desk for days on end would have been even more isolating that it already was had I not been able to vent, talk about the struggle, and celebrate small victories. Having a cheerleading section who also is intimately familiar with how hard it can be was crucial in completing the PhD. Bonnie Bui and Stephanie Pulles reminded me that this process, with all its ups and downs, is
universal. It was funny and comforting how even being across the country in completely different fields, Bonnie and I were somehow going through the same things. Karen Duek was my first true friend in the program. I am grateful that she reminded me to have compassion for myself during this process. At times, and not nearly enough, I was able to retreat to Connecticut or New Hampshire and stay with Kate and Paul Dawson. These visits were always restorative and reminded me that I had to try and find a way to maintain some semblance of a work-life balance. Kate would always remind me to take a break every once in a while, which was much needed but often unheeded advice, and that somatic effects of stress were not something to just endure. Stephanie Kakadelis was one of the first people I would vent to on Google hangouts. Thanks for providing an outlet for my frustration, especially when I was writing sections I did not like. I feel lucky that I had a close friend, Dan Mann, who decided around the same time to respecialize and overcome the inertia of staying in a subfield that was not working out. It was beneficial to be able to talk about that and not feel like the only one to respecialize mid-PhD. Lastly, weekly sushi, drinks, and RuPaul viewing with Jorge Baquero and Carlos Devia invigorated me. Jorge and Carlos modeled commitment with balance, sometimes even telling me that I couldn’t hand out or go dancing because I had to write. They importantly provided a connection to my community that was missing being so far away from family. Carlos pushed me to always be moving forward and sometimes just get it done. Chats lubricated with vodka soda (with a splash of cranberry) or wine would snap me out of any funk I was in. I am especially grateful to Jorge for going over this dissertation and making sure my heritage Spanish was fit for print.

I am grateful to have friends scattered across the country that supported me in this hustle. I was able to take a moment and enjoy myself with them. I am especially appreciative for my dear friends’ patience with me over the years when I would go MIA for months on end. I feel really lucky that you continued to be my friend even when I would not answer a text message for a month. Thank you Josh Heckroth, Carlos Becerril, Jocelyn Loren, Lisa Kimsey, Kai Tuft, and Sarah Brady Stack. I am lucky to have Candace Davider in my life for all these years. She has been my biggest cheerleader and even was the person to suggest I major in Linguistics. On her advice, when I resumed my studies I declared my major to be Linguistics without knowing what it was. She has been cheering me on since community college and even when I was a college drop out. I am immeasurably blessed to have a friend, and former
babysitter, that dreamed bigger for me that I would ever have hoped. I am grateful to my buddy, Jay R Abeyta, for being the friend that was always honest when I’d wonder if I’d ever finish. Sometimes the answer was “No, unless you stop procrastinating” and I’m grateful for the honesty. Just like in undergrad when we were roommates and for my MA thesis, Jay helped me with my formatting and bibliography because I stubbornly never wanted to put in the time in these 11 years to learn Zotero. Thank you for the manual references management. IN addition to helping me with the dissertation, Jay generously opened his home to me when I needed to get out of NYC and just hole up and work. I’m grateful he didn’t think it was rude when I’d visit very 6-9 months and only leave the room to get more diet coke. I am lucky to have more than one home outside of the city that I could go to work, relax, and be with an old friend, Meghan Tucker. I am grateful to have a friend like Meghan that understands me, supports me, and didn’t stop calling me. After the defense, it was hard to find the motivation to complete revisions. I’m appreciative of the enthusiastic encouragement that my boyfriend, Samy el-Noury, gave me to jump over the last hurdle. Lastly, I am grateful to the men and women I met doing bowling, kickball and dodgeball with Big Apple Rec. It provided a much needed outlet away from work.

Without my therapist, Eliezer Mayer, I would definitely have either quit the program or this would have taken much much longer. I am a better person for the 5 years I’ve been seeing him. I am very proud of this dissertation in part because I feel reflected in the work. This was only possible by working with him.

In addition to the unwavering support of my friends, my family’s belief in me helped me feel like this task was doable. My family was so thoroughly convinced I could do this that there was no doubt. I am grateful to my family who openly and frequently expresses love, appreciation, compliments, and support. The majority of my family is across the country, so I am grateful that my White cousins, Tiffany and Josh Smith, were nearby. I am grateful for their support and long talks at the dinner table. This study of Latinx Spanish speakers in the anglophone US is deeply personal. I am grateful that I was able to complete work on an understudied minority group, a group of which I and my family are members. I thought about my family constantly when writing and it spurred me on to finish the dissertation even when I just wanted to walk away. Thank you Sylvia and Christopher Livingston and Sandra Serrano Hernandez for the
unconditional love. And lastly but most importantly, my Granma, Bertha García Serrano, would remind me of how fortunate I am to have the opportunity to study as much as I wanted. She loved school and loves learning but her family pulled her out of school in Mexico in the 4th grade and her husband wouldn't let her take English classes when she immigrated to California. So, in spite of all the obstacles, I was driven to earn a PhD. I am grateful for the love and support of my Granma.
# TABLE OF CONTENTS

**ABSTRACT** ................................................................................................................................. iv  
**ACKNOWLEDGEMENTS** ............................................................................................................. vi  
**TABLE OF CONTENTS** ............................................................................................................... xiii  
**TABLE OF TABLES** ..................................................................................................................... xviii  
**TABLE OF FIGURES** ................................................................................................................... xx  
**LIST OF ABBREVIATIONS** .......................................................................................................... xxii  

## CHAPTER I: INTRODUCTION

1. Heritage speaker processing .................................................................................................. 1  
2. Aims of the proposed study ................................................................................................. 2  
3. Importance of study ............................................................................................................. 2  
4. Outline of dissertation .......................................................................................................... 3  

## CHAPTER II: HERITAGE SPEAKERS

1. Heritage Speakers ............................................................................................................... 4  
   1.1 Definition of heritage speakers ..................................................................................... 5  
   1.2 Prevalence of heritage speakers ................................................................................... 7  
2. Spanish heritage speakers in US ......................................................................................... 8  
3. Empirical profile of HS language exposure and use ....................................................... 10  
   3.1 Methodology .................................................................................................................. 12  
      3.1.1 Participants .............................................................................................................. 12  
      3.1.2 Protocol .................................................................................................................. 12  
      3.1.3 Variables of interest ............................................................................................... 13  
   3.2 Principal component analysis ....................................................................................... 14  
      3.2.1 Rotated components ............................................................................................... 15  
      3.2.2 Group comparisons ............................................................................................... 18  
   3.3 Discussion ....................................................................................................................... 21  
4. Present issues with the studies of heritage speakers ...................................................... 22  
   4.1 Prescriptivist and monolingual comparison ................................................................. 24  
   4.2 Linguistic insecurity ....................................................................................................... 25  
   4.3 Empirical study of linguistic insecurity ....................................................................... 28  
      4.3.1 Methodology ............................................................................................................ 28  
         4.3.1.1 Participants ....................................................................................................... 28  
         4.3.1.2 Stimuli ............................................................................................................... 29  
         4.3.1.3 Protocol ............................................................................................................. 31  
         4.3.1.4 Data analysis ..................................................................................................... 32  
   4.3.2 Results ....................................................................................................................... 33
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2 Stimuli</td>
<td>77</td>
</tr>
<tr>
<td>3.3 Experimental design &amp; procedures</td>
<td>78</td>
</tr>
<tr>
<td>3.4 Data acquisition</td>
<td>79</td>
</tr>
<tr>
<td>3.5 Data processing</td>
<td>79</td>
</tr>
<tr>
<td>4. Voltage data</td>
<td>80</td>
</tr>
<tr>
<td>5. Data analysis</td>
<td>84</td>
</tr>
<tr>
<td>5.1 Temporo-spatial analysis</td>
<td>86</td>
</tr>
<tr>
<td>5.1.1 Temporal analysis</td>
<td>87</td>
</tr>
<tr>
<td>5.1.2 Spatial analyses</td>
<td>89</td>
</tr>
<tr>
<td>5.1.3 Results</td>
<td>91</td>
</tr>
<tr>
<td>5.2 Linear mixed-effects modeling</td>
<td>92</td>
</tr>
<tr>
<td>5.3 Across ERP component analysis</td>
<td>94</td>
</tr>
<tr>
<td>6. Results</td>
<td>95</td>
</tr>
<tr>
<td>6.1 Frontal P600</td>
<td>95</td>
</tr>
<tr>
<td>6.2 Centroparietal P600</td>
<td>99</td>
</tr>
<tr>
<td>6.3 N420</td>
<td>104</td>
</tr>
<tr>
<td>6.3 N375</td>
<td>108</td>
</tr>
<tr>
<td>6.5 Across ERP components</td>
<td>111</td>
</tr>
<tr>
<td>7. Discussion</td>
<td>115</td>
</tr>
<tr>
<td>7.1 Comparison to previous studies</td>
<td>118</td>
</tr>
<tr>
<td>7.2 Temporo-spatial PCA</td>
<td>118</td>
</tr>
<tr>
<td>7.3 Lack of P600-effect for heritage speakers</td>
<td>121</td>
</tr>
<tr>
<td>7.4 Opposing N400-like effects</td>
<td>122</td>
</tr>
<tr>
<td>7.5 Heritage speaker processing difficulty</td>
<td>124</td>
</tr>
<tr>
<td>7.6 Future directions</td>
<td>127</td>
</tr>
<tr>
<td>8. Summary</td>
<td>128</td>
</tr>
</tbody>
</table>

CHAPTER V: TASK-EVOKED PUPILLARY RESPONSES                                  129

1. Pupillometry and language processing.                                    129
   1.1 Pupillometric studies of language                                     129
   1.2 Pupillometric studies of relative clause processing                  131
   1.3 Pupillometric studies of multilingual processing                    133

2. Research question & hypothesis                                           135

3. Methods                                                                136
   3.1 Participants                                                        136
   3.2 Ancillary Measures                                                  137
      3.2.1 Working memory                                                  137
      3.2.2 Spanish Comprehension                                           138
      3.2.3 Spanish Oral Fluency                                            138
TABLE OF TABLES

Table 1: Factor loading by language background questionnaire items .......................................................... 17
Table 2: Mean component scores for work, media, and English use factors by bilingual group. .......... 19
Table 3: Mean component scores for English and Spanish identity factors by bilingual group. .......... 20
Table 4: Mean component scores for English and Spanish ability factors by bilingual group. .......... 21
Table 5: Mean participant-average naturalness rating by complex noun phrase island condition and by- and across- bilingual group ........................................................................................................ 34
Table 6: Mean participant-average naturalness rating by relative clause island condition and by- and across- bilingual group ........................................................................................................ 34
Table 7: Mean participant-average naturalness rating by temporal adverbial island condition and by- and across- bilingual group ........................................................................................................ 35
Table 8: Mean participant-average naturalness rating for complement clause islands and across all condition and by- and across- bilingual group. ................................................................. 35
Table 9: Mean participant-average naturalness rating by wh-island condition and by- and across- bilingual group .................................................................................................................... 35
Table 10: Distribution of naturalness ratings by- and across- group by grammaticality. ....................... 36
Table 11: Maximal zero-and-one inflated beta regression mixed-effects model coefficients by rating distribution parameters. ........................................................................................................... 37
Table 12: Beta regression estimates and geometric rating means by group and by condition. .......... 38
Table 13: Beta regression estimates and odds ratio by group and by condition. ................................. 38
Table 14: Experimental evidence of subject-object relative clause processing asymmetry. ............ 55
Table 15: European crosslinguistic evidence of subject-object relative clause processing asymmetry. ... 56
Table 16: Non-European crosslinguistic evidence of subject-object relative clause processing asymmetry. ................................................................. 57
Table 17: Mean participant-average accuracy (proportion correct) by bilingual group and by condition... 65
Table 18: Non-significant maximal logistic mixed-effects model of accuracy. .................................................. 66
Table 19: Mean frontal P600 amplitude by bilingual group and condition. .................................................. 97
Table 20: Most parsimonious linear mixed-effects model of frontal P600 amplitude............................. 98
Table 21: Maximal linear mixed-effects model of frontal P600 amplitude ....................................................... 98
Table 22: Confidence intervals for model parameters of frontal P600 amplitude ....................................... 99
Table 23: Mean centroparietal P600 amplitude by bilingual group and condition. .................................... 101
Table 24: Maximal linear mixed-effects model of centroparietal P600 amplitude ..................................... 102
Table 25: Predicted marginal means from maximal model of centroparietal P600 amplitude ................ 103
Table 26: Mean N420 amplitude by bilingual group and condition. ............................................................. 104
Table 27: Maximal linear mixed-effects model of N420 amplitude ............................................................. 106
Table 28: Predicted marginal means from maximal model of N420 amplitude ........................................ 106
Table 29: Mean N375 amplitude by bilingual group and condition. ............................................................. 108
Table 30: Maximal linear mixed-effects model of N375 amplitude ............................................................. 110
Table 31: Predicted marginal means from maximal model of N375 amplitude ................................................................. 110
Table 32: ERP component correlations and regression coefficients with 95% confidence intervals across bilingual group .............................................................................................................. 111
Table 33: ERP component correlations and regression coefficients with 95% confidence intervals for late bilinguals ...................................................................................................................... 113
Table 34: ERP component correlations and regression coefficients with 95% confidence intervals for heritage speakers ........................................................................................................... 114
Table 35: ERP component results ........................................................................................................................................ 116
Table 36: Mean percent of English use for language exposure variables .................................................................................. 125
Table 37: Mean percent of Spanish use for current language use variables .............................................................................. 125
Table 38: Mean percent of language ability by language by skill .................................................................................................. 127
Table 39: Mean trial-average dilation by bilingual group and condition ................................................................................... 145
Table 40: Maximal linear mixed-effects model of trial-average dilation ..................................................................................... 147
Table 41: Predicted marginal means from maximal model of trial-average dilation ................................................................. 147
Table 42: Mean trial maximum dilation by bilingual group and condition .................................................................................. 149
Table 43: Maximal linear mixed-effects model of trial maximum dilation .................................................................................. 151
Table 44: Predicted marginal means from maximal model of trial maximum dilation ................................................................. 151
Table 45: Mean epoch-average dilation by bilingual group and condition .................................................................................. 152
Table 46: Maximal linear mixed-effects model of epoch-average dilation .................................................................................. 154
Table 47: Predicted marginal means from maximal model of epoch-average dilation .............................................................. 154
Table 48: Significance of models by 500ms window. .................................................................................................................... 156
Table 49: Mean average dilation in window 7 (3001-3500ms) by bilingual group and condition .............................................. 158
Table 50: Maximal linear mixed-effects model of average dilation in window 7 ........................................................................ 158
Table 51: Predicted marginal means from maximal model of average dilation in window 7 ...................................................... 159
Table 52: Dependent variable correlations across bilingual group. ............................................................................................ 160
Table 53: Dependent variable correlations for late bilinguals. ...................................................................................................... 161
Table 54: Dependent variable correlations for heritage speakers. ............................................................................................... 162
Table 55: TEPR results ............................................................................................................................................................... 163
Table 56: Mean percent of English use for language exposure variables .................................................................................. 167
Table 57: Mean percent of Spanish use for current language use variables .............................................................................. 167
Table 58: Mean percent of language ability by language by skill .................................................................................................. 169
Table 59: Two-tailed empirical p-value and log-transformed p-value of the observed summed difference for late bilingual participants from empirical frequency distribution for significant effects .......................................................... 181
Table 60: Two-tailed empirical p-value and log-transformed p-value of the observed summed difference for late bilingual participants from empirical frequency distribution for significant effects .......................................................... 182
# TABLE OF FIGURES

Figure 1: Correlation matrix of language use, behavior, and exposure variables across group ................................................................. 15
Figure 2: Factor loadings by language background questionnaire items by component .......................................................... 16
Figure 3: Mean component scores for work, media, and English use factors by bilingual group ................................................................. 18
Figure 4: Mean component scores for English and Spanish identity factors by bilingual group ................................................................. 20
Figure 5: Mean component scores for English and Spanish ability factors by bilingual group ................................................................. 21
Figure 6: Mean participant-average naturalness ratings by condition by grammaticality by bilingual group. .............................. 34
Figure 7: Languages showing evidence of subject-object processing asymmetry ................................................................. 55
Figure 8: Spanish RMST object-relative clause item picture array ........................................................................................................ 63
Figure 9: Spanish RMST relative clause participant-average mean accuracy (proportion correct) across condition and bilingual group ........................................................................................................ 64
Figure 10: Grand mean waveforms for subject-relatives (solid, blue line) and object-relatives (dashed, red line) by-electrode by-millisecond ................................................................. 81
Figure 11: Grand mean waveforms for subtraction waveforms by-electrode by millisecond ................................................................. 81
Figure 12: Grand mean waveforms by individual left and central electrode by millisecond ................................................................. 82
Figure 13: Grand mean waveforms by individual posterior electrode by millisecond ................................................................. 83
Figure 14: Grand mean waveforms by individual left anterior electrode by millisecond ................................................................. 84
Figure 15: General ERP data analysis flowchart .............................................................................................................................. 85
Figure 16: Proposed temporo-spatial PCA ERP data analysis flowchart .............................................................................................. 86
Figure 17: Factor weights by millisecond for the first rotated temporal component ................................................................. 88
Figure 18: Factor weights by millisecond for the seventh and nineteenth rotated temporal component ................................................................. 89
Figure 19: Results of the temporo-spatial PCA ERP data analysis flowchart ................................................................. 91
Figure 20: Model selection ................................................................................................................................................................. 92
Figure 21: Group-condition contrasts from linear mixed-effects model and from contrasts from confidence intervals for predicted marginal means from most parsimonious mixed-effects model. ................................. 94
Figure 22: Scalp array of scaled amplitude of the frontal P600 ......................................................................................................................................................... 96
Figure 23: Grand mean waveform of frontal P600 amplitude by condition ............................................................................................... 96
Figure 24: Mean frontal P600 amplitude by bilingual group and condition ............................................................................................... 96
Figure 25: Linear mixed-effects models for frontal P600 amplitude ........................................................................................................... 98
Figure 26: Scalp array of scaled amplitude of the centroparietal P600 ........................................................................................................ 99
Figure 27: Grand mean waveform of centroparietal P600 amplitude by condition ................................................................................... 100
Figure 28: Mean centroparietal P600 amplitude by bilingual group and condition ................................................................................... 100
Figure 29: Linear mixed-effects models for centroparietal P600 amplitude .............................................................................................. 102
Figure 30: Scalp array of scaled amplitude of the N420 ......................................................................................................................................................... 103
Figure 31: Grand mean waveform of N420 amplitude by condition ............................................................................................... 104
Figure 32: Mean N420 amplitude by bilingual group and condition ............................................................................................... 104
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
</table>
| EEG          | Electroencephalograms |}
| ERP          | Event-related potentials |}
| fMRI         | Functional magnetic resonance imaging |}
| HL           | Heritage language |}
| L1           | First-learnt language |}
| L2           | Second-learnt language or non-First-learnt language |}
| LOLT         | Language of learning and teaching |}
| ML           | (Social) majority language |}
| PCA          | Principal component analysis |}
| RMST         | RISLUS multilingual syntax test |}
| TEPR         | Task-evoked pupillary responses |}
CHAPTER I:
INTRODUCTION

1. Heritage speaker processing

Models of grammar, processing and acquisition are primarily built on evidence from monolinguals and adult learners of a second language. Heritage speakers, who are bilinguals of a societal minority language, acquire and use their heritage language in informal settings; but who live, work, and are educated in the societal majority language. The differences between heritage speakers and both monolinguals and adult second language learners are extensive: heritage speakers are not educated in the heritage language, their input is typically not from a prestige variety of the heritage language, and they are dominant in the majority language, using it more frequently (Valdés, 1989).

Research into the acquisition and linguistic competence of these heritage speakers extends the same assumptions and well-developed and researched methodologies from second language acquisition research. This results in a number of inappropriate assumptions. First, monolingual controls are an inappropriate comparison group because they learn and use language very differently from heritage speakers. Second, there is an inappropriate focus on features of the prestige variety of the heritage language which are typically acquired, reinforced, and normalized through standardized education (Flores, 2014). Third, there is an inappropriately widespread use of metalinguistic tasks, which require reflection on grammaticality or acceptability in a population noted for its linguistic insecurity (Klein & Martohardjono, 2008). Consequently, heritage speaker grammars are characterized as simple, decayed/attributed, and incomplete (Benmamoun, Montrul, & Polinsky, 2010; Scontras, Fuchs, & Polinsky, 2015), and are compared to intermediate second language learner grammars (Montrul, 2005). The predominant deficit view of the heritage speaker grammar is further confounded not only by inappropriate comparisons and inappropriate tasks, but also by the lack of adequate models of heritage speaker processing.
2. Aims of the proposed study

The present study addresses these shortcomings by investigating heritage speaker language processing in the heritage language using appropriate measures of an appropriate linguistic phenomenon compared to appropriate baseline group. The first aim of the study is to motivate and utilize an appropriate group for baseline comparisons. Heritage speakers are therefore compared to adult late second language learners living and working in a second language dominant society from the same community. The study focuses on fluent Spanish and English Latinx\(^1\) bilinguals living in the anglophone US. Spanish heritage speakers are appropriately compared to their time-apparent parents (English speaking Latinx immigrants who moved to the anglophone US in adulthood). Latinx use of and abilities in English and Spanish are analyzed to establish a profile of language exposure, use, ability, and identity. The identified patterns are then explored across group to establish ways that heritage speakers and late bilinguals are similar and different. The second aim of the study is measure language processing by utilizing an appropriate linguistic phenomenon. We therefore investigate the subject-object relative clause processing asymmetry in heritage speakers. The asymmetry is a robust phenomenon present cross-linguistically utilizing various research methodologies. It has been well-established in both Spanish and English, is early acquired, and is not confounded by prescriptive rules or literacy. The third aim of the study is to utilize an appropriate measure of processing. Heritage speakers are therefore tested using pupillometry and event-relative potentials. Pupillometry and event-relative potentials are non-invasive, implicit measures that allow for auditory presentation of Spanish sentences. The protocols measure automatic physiological responses to neural language processing in real-time and do not require metalinguistic judgments.

3. Importance of study

This study contributes an important first step in developing best practices for research of heritage speaker grammar and language processing. The current study demonstrates that language dominance, which results in heritage language anxiety, is an important and overlooked confound that has implications in

\(^{1}\) We use the term “Latinx” as a non-binary, inclusive term as opposed to the male-gendered “Latino”, “Latino/a”, or “Latin@”. In this dissertation, the term “Latinx” can be considered to be interchangeable with “Hispanic” as the population we are focusing on are “Latinx Hispanics” to the exclusion of Hispanic, non-Latinx Spaniards or Equatoguinean; and non-Hispanic, Latinx Brazilians, Surinamese, Guyanese, and Guianan.
properly measuring processing and theorizing about the grammar of heritage speakers. The findings allow us to make testable predictions and clear next steps in developing a research protocol that is appropriate for heritage speakers. The study clarifies the importance of operationalizing language dominance to avoid prevalent confounds in heritage speaker studies.

4. Outline of dissertation

This dissertation is laid out as follows. Chapter two discusses heritage speakers and identifies confounds in the study of their language processing and grammar. A principal component analysis and group difference testing is run to identify similarities and differences between heritage speakers and late bilinguals in their language use, exposure, ability, and identity. The major confound of linguistic insecurity is introduced and a study of metalinguistic task performance by heritage speakers and late bilinguals is presented and interpreted. Chapter two, lastly, critically reviews the predominant models of bilingual language processing. Chapter three presents the rationale for the present study and motivates the implicit measures used. The appropriate linguistic phenomenon used in the study is introduced. Evidence of heritage speaker competence in relative clause comprehension is established. Chapter four reviews the relative clause processing literature and bilingual studies utilizing the event-related potential methodology. Data from an event-related potential study are analyzed using a temporo-spatial analysis to identify four ERP components. Late bilingual and heritage speaker ERP component results are presented and discussed. Chapter five reviews the relative clause processing literature and bilingual studies utilizing the pupillometry methodology. Data from a pupillometry study are analyzed as four task-evoked pupillary response measures. Late bilingual and heritage speaker task-evoked pupillary response results are presented and discussed. Chapter six concludes with a summary of the findings from the five studies in previous chapters. A combinatoric study comparing event-related potential and pupillometry is run. The findings of all six studies and the influence of language dominance are discussed. A model of heritage speaker language processing utilizing more efficient processing strategies to account for confounds is proposed. Recommendations for experimental linguistic research and future studies are presented.
CHAPTER II:
HERITAGE SPEAKERS

Research in linguistics has principally focused on the acquisition, processing and grammatical knowledge of monolingual speaker/listeners. Second language acquisition research has principally focused on the acquisition of a second language by adults. The models and theories from these research programs have given insight into the cognitive processes underlying language; however, their limited scope overlooks the large proportion of multilingual speakers worldwide. Heritage speakers, first language speakers of a socially undervalued minority language, are prevalent. In the anglophone US alone, heritage speakers represent 20% of people under the age of 18 (Ryan, 2013), yet studies researching their unique acquisition, processing and grammatical knowledge are limited in number.

In this chapter, heritage speakers will be discussed along with the potential confounds that arise in the study of their grammatical knowledge and language processing. Section one will introduce heritage speakers by explaining what type of bilingual they are, contextualizing heritage speakers in terms of prevalence, and then narrowing the scope to US Latinx heritage speakers, who are the focus of the present study. A principal component analysis of items from a questionnaire administered to US Latinx heritage speakers, including language use, exposure, ability, and identity is conducted. The resulting variables are analyzed and group similarities and differences in late bilinguals (heritage speakers’ apparent-time parents) and heritage speakers are presented. Linguistic insecurity in metalinguistic tasks is reviewed and a study of metalinguistic task performance patterns in heritage speakers and late bilinguals is presented. Section two will motivate the focus on language processing in heritage speakers by discussing issues with research focused on underlying representations of the grammar. Finally, the major models of bilingual language processing and their applicability to heritage speakers will be presented.

1. Heritage Speakers
This section defines heritage speakers and describes the debate surrounding them. Deficit-based approaches to studying heritage speaker grammar such as theories of incomplete acquisition and attrition
are introduced (Benmamoun, Montrul, & Polinsky, 2010) along with problems regarding those approaches. Bilinguals outnumber monolinguals, and although heritage speakers are a group whose heritage language is rarely studied because the heritage language is devalued, and speakers are generally not educated in the heritage language, they represent a huge number worldwide. Understanding heritage language knowledge will benefit the field of linguistics as well as provided needed information to education professionals who have up to this point concentrated on heritage language speakers’ deficit in the societally dominant language and not their skills in the home language. The representation of heritage speakers with a home language of Spanish in the anglophone US and New York City is quantified, as New York City heritage speakers of Spanish are the population that that this dissertation investigates. An analysis of language exposure and use differences in heritage speakers is presented. Critical issues in the study of heritage speakers are explored, including the issues with previous studies comparing heritage speakers to prescriptivist norms and monolingual control groups. Empirical proof of linguistic insecurity is introduced. The section on heritage speakers concludes with an argument of the need for more studies of heritage speakers generally, and specifically the need for a new way to study these speakers that is different from the monolingual controls, prescriptive variety comparisons, and metalinguistic judgment tasks that makes up most of the heritage speaker research up to this point.

1.1 Definition of heritage speakers

There is no single standard definition of a heritage speaker. Definitions of heritage speakers vary within the field of linguistics, and have even greater differences when compared to other fields such as education. Researchers are then each left the task of defining the heritage speaker population, which makes it difficult to compare heritage speakers across studies. Much of this disorganization is a result of the complex web that characterizes the field of bilingualism in general, and so that web is herein disentangled. Bilinguals (in this dissertation, defined as individuals who use two or more languages in their daily life) are typically characterized according to proficiency and order of language acquisition. Proficiency can be characterized in relation to “nativeness” (e.g. as native, near-native, or learner), or in terms of relative ability regarding their languages (e.g. balanced bilinguals vs. dominant bilinguals). Order
of acquisition can be as specific as specifying the age of onset for each language, or as general as first language (L1) vs. second language (L2). Individuals who learn one language after another are called sequential bilinguals, who are divided by Montrul (2008) into the groups of early sequential bilinguals, who begin to learn the second language between ages four and twelve, and late sequential bilinguals, who begin to acquire the second language after age twelve. Simultaneous bilinguals are those who learn both languages from birth and therefore have two L1s.

Heritage speakers are broadly defined as individuals who experience a shift in dominance from the heritage language (HL) to the societal majority language (ML) at a young age, typically before age eight (Silva, 2008). Note that the distinction in language dominance is discussed in terms of HL and ML, and not in terms of first-learnt language (L1) and second-learnt language (L2). This is because the heterogeneity of heritage speakers, being either sequential or simultaneous bilinguals, means that these terms are not interchangeable. A heritage speaker may have two L1s and therefore cannot shift in dominance in traditional notions.

Heritage speakers do not fit neatly into any of the slots generally used to define bilinguals. In terms of proficiency, heritage speakers range from balanced bilinguals who have a strong command of the prestige variety and written modality in both languages, to individuals who may understand the home language but who may not be comfortable speaking the home language (called receptive bilinguals). An expanded definition of heritage speakers includes those with a family connection to the language (Fishman, 2001; Beaudrie & Fairclough, 2012), regardless of proficiency. For example, an American of Armenian descent who only knows a limited number of Armenian words or phrases whose family, as a consequence of the family’s failure to use or transmit Armenian due to assimilation pressures, is also classified as an Armenian heritage speaker under a broader definition. The present study limits the operational definition of heritage speakers to speakers that are proficient communicators in their home or family language. This is similar to Duarte (2014), who limits the operational definition of heritage speakers to bilinguals “whose level of proficiency in speaking allows [them] to interact with other proficient speakers of the language” (p. 38).

In terms of nativeness, heritage speakers may identify themselves as a native speaker of the heritage language, or the societal majority language, or both (or neither). As mentioned above, even the
labels L1 and L2 are debatable, as heritage speakers can be simultaneous bilinguals or early sequential bilinguals. What researchers across fields do tend to agree on is that heritage speakers are a heterogeneous group with a wide range of abilities in both of their languages (Zhang, 2017).

Heritage speakers, growing up in a society wherein the majority of people speak and use another language, are typically educated in a ML-medium classroom. Explicit instruction in their home language is limited to foreign language classes in high school or college. Similarly, they might not “formally learn” the language and its prescriptivist grammatical rules and formal constructions unless they attend afterschool or weekend programs (e.g., evening Mandarin classes or weekend Greek school). For example, in New York City, heritage speakers of Spanish or Arabic are typically educated only in English-medium classrooms as English is the default language of learning and teaching (LoLT) from pre-school through college (Beaudrie & Fairclough, 2012; Menken & Kleyn, 2010; Valdés, Fishman, Chavez & Perez, 2006). In spite of using the HL regularly, they might only learn the formal variety in college classes of be exposed to it through media. Some heritage speakers may indeed not even be literate in their home language. However, ML literacy skills can transfer resulting in functional literacy in the HL (Duarte, 2014). The result of this situation is that heritage speakers have vastly different backgrounds in terms of literacy and education compared to monolinguals.

1.2 Prevalence of heritage speakers

Unfortunately, there are no reliable estimates of multilingualism worldwide making it difficult to estimate the number of heritage speakers. However, the ubiquity of multilingualism suggests that heritage speakers are more common than one would assume given the focus on monolinguals and late bilinguals in experimental language research. Multilingualism is increasingly becoming the global norm and is currently widespread, especially among economically and politically disadvantaged group (Myers-Scotton, 2006). There are nearly 1.5 billion L2 speakers worldwide of only four languages (English: ~750 million; Hindi: ~275 million; French: ~210 million; Mandarin Chinese: ~200 million; Ethnologue, 2018), likely a result of these languages being the primary lingua franca of the global economy and the lingua franca of former colonial multietnic/multilingual nation states. The prestige given these languages in business and education results in limited schooling in the home languages of people worldwide.
Multilingualism is widespread in former colonized regions, such as West Africa (75% of people in Guinea-Bissau and 75% in Niger speak more than one language; Hovens, 2002) and India (nearly 1-in-3 people speak two or more languages; Census India, 2001). In the Global North, multilingualism is also common. In the European Union, more than half (54%) of people can have a conversation in more two or more languages (European Commission, 2012).

In the anglophone US\(^2\), more than one-in-five (21%; \(n=60.6\) million) people ages 5 and older speak a language other than English in the home, which is a 163% increase from 1980 (Ryan, 2013). Given the near universality of English-only education in the anglophone United States, it is reasonable to estimate that there are 13.8 million heritage speakers in the United States, representing 20% of children 5-18 years old. (numbers taken from Ryan, 2013). Multilingualism and the prevalence of heritage speakers will likely only increase with the current refugee crisis and migration to the global North as more than half of the more than 65 million refugees worldwide are school aged children (Park, Katsiaficas, & McHugh, 2018).

2. Spanish heritage speakers in US

In the US, Spanish speakers are by far the largest population of heritage speakers. The US is frequently characterized as an English-speaking country and Spanish speakers are made to feel un-American for speaking Spanish. However, Spanish speakers in the anglophone US are not all immigrants or from recent immigrant families. The oldest continuous European settlement in the contiguous US, St. Augustine, Florida, was founded by the Spanish empire. As a consequence of Manifest Destiny imperialism and colonialism, 49% of the contiguous US was once a part of the Spanish Empire. The Chicanx saying, “Nosotros no cruzamos la frontera, ésta nos cruzó a nosotros” / “We didn’t cross the border, the border crossed us”, referring to the American invasion and colonialization of the northern

---

\(^2\) Throughout this dissertation, we will use the term “anglophone US” to refer to the English-dominant territories of the United States of America. English is the de facto language of government and education in the 50 states and the District of Columbia, the national capital territory. Puerto Rico, a territory of the US, is a Spanish-dominant society with Spanish being co-official with English in government and Spanish being the primary language of learning and teaching. As the term “US” necessarily includes Puerto Rico a distinction is made. We use the term “US” to refer to all of the multilingual United States of America, including regions like Spanish-dominant Puerto Rico, Samoan-dominant American Samoa, Chamorro-minority Guam and Northern Mariana Islands, and any of the varyingly multilingual 573+ indigenous nations.
Mexican states, highlights the fact that multilingualism in the anglophone US is not new, but rather a persistent and core characteristic of the US. In fact, the US is the second largest Spanish-speaking country in the world, after Mexico. Spanish is the language most widely spoken after English with 61.6 million L1 and L2 Spanish speakers in the anglophone US and Puerto Rico (Ethnologue, 2018; Moreno Fernández & Otero Roth, 2006). In the greater New York City area, there are 3.5 million Spanish speakers that make up 19.7% of the population (Ryan, 2013). Unlike other languages spoken in the US, such as Italian, Danish, or Polish, the continuing migration north for political, humanitarian, and economic reasons of Spanish-speaking Northern, Central, and Southern American individuals, helps to actively maintain Spanish as an American language. In the US, the Spanish language is interwoven with the ethno-racial identity of the US Latinx community. While not all Latinxs speak Spanish and not all Spanish speakers are Latinx, between 75%-80% of Latinxs speak Spanish (Ryan, 2013; Krogstad & Lopez, 2017). Bilingualism is encouraged in US Latinxs and 88% of Latinxs feel it is important that future generations are able to speak Spanish (Lopez, Krogstad, & Flores, 2018).

While it is difficult to accurately estimate the number of Spanish heritage speakers, nationwide there are 9.9 million Spanish-speakers aged 5-18 years old that represent 15.7% of school-aged children that could be classified as heritage speakers when they are adults. The New York City Department of Education (2017) reports that out of the 150,741 students receiving services for English language learners in the 2015-16 school year, 92,746 (61.53%) of them speak Spanish. Additionally, 15,270 Spanish-speaking English language learners are characterized as long-term English language learners, who have received English language support for eight years or more, but have still not reached the score of proficient on the New York State English as a Second Language Achievement Test.

Negative attitudes by the anglophone majority serve to discourage bilingualism and heritage language use (Fishman, 1991), and in the field of education, the English-only movement and legislation prohibiting the use of heritage languages in schools has had the effect of accelerating language loss among families who speak languages other than English (See Baker, 2006 and Crawford, 2008 for a review of the English-only movement and its impact on education legislation). These pressures are not exclusively external to the Spanish-speaking community. Twenty percent of Latinxs in the US report that their parents discouraged them from using Spanish when growing up (Lopez, Krogstad, & Flores, 2018).
Decline of Spanish use is typically seen across immigrant generations. First generation parents are more likely to encourage the use of Spanish than third or later generations (Lopez, Krogstad, & Flores, 2018). Likewise, there is a decline in regular use of Spanish or Spanish and English from 44% among the second generation to 24% in the fourth or later generations (Krogstad & Gonzalez-Barrera, 2015).

3. Empirical profile of HS language exposure and use

Heritage speakers, being raised in a multilingual and ML dominant environment, use Spanish and English differently than their parents and also differently from monolingual comparisons in HL dominant societies. Heritage speakers are educated in ML-medium classrooms and their peers and family are fluent in the ML. Heritage speakers use the HL most commonly with their parents (Garcia & Diaz, 1992; Nguyen, Sin & Krashen, 2001) and older relatives (Carreira, Jensen, & Kagan, 2009). The HL is used less with their siblings (Garcia & Diaz, 1992; Nguyen, Sin & Krashen, 2001) and younger relatives and children (Galindo, 1996).

Interestingly, language use takes on a specialized function in heritage speakers with the HL and ML being associated with different domains, activities, or subject matter such as the ML for school-related matters and the HL for religious matters. The alternation of languages serves an additional communicative purpose (Delgado, 2009) not available to monolinguals to whom heritage speakers are so often compared. The HL is used most commonly in the private domains of the house and the ML is used in public domains such as work and in school (Galindo, 1996). That is not to say that the HL is not used in public spheres; Heritage speakers may use the HL when interacting with other members of their ethno-lingual group to reinforce in-group status (Phinney et al., 2001), or index levels of intimacy (Delgado, 2009), and to talk discretely in public (Carreira, Jensen, & Kagan, 2009). Heritage speakers have also been claimed to select the HL and ML in order to accommodate the dominant language of the interlocutor (Delgado, 2009).

Although raised in an ML dominant society, among Spanish heritage speakers in the anglophone US, all report listening to Spanish language music and 70% report watching Spanish language TV at least once a month. Most commonly (32%) to watch telenovelas, prime-time single run television drama series (Velázquez, 2015). Interestingly, although not educated in Spanish or explicitly taught Spanish literacy
skills, US Spanish heritage speakers report high rates of reading and writing in Spanish due to social media and SMS texting use (Velázquez, 2015).

Extra-linguistic factors impact the use of the HL versus ML in heritage speakers in a complex way because heritage speakers typically are members of a low-prestige ethnic-minority group. Language and ethnic identity are closely linked in a way that one doesn't find in adult L2 learners of a language in a foreign language setting. The White American adult L2 learner of Spanish has an ethnic identity as a White American that is not linked to their English proficiency although they might use Spanish with Latinx individuals as a way of distancing themselves from their ethnicity. A heritage speaker’s language use profile is therefore not stable across their lifetime, but closely tied to ethnic identity. Tse (1998) argues that HL language use is lowest in late childhood and adolescence due to attitudes of ethnic ambivalence/evasion. Interest in the HL is renewed in adulthood, if at all. Consequently, HL proficiency is closely related to ethnic identity (Phinney et al., 2001).

To explore the language use, exposure, ability, and identity of heritage speakers with respect to their HL and ML, we analyze a language background questionnaire administered to Spanish heritage speakers in the anglophone US. The patterns of language use discussed in previous literature typically present the findings of only heritage speakers, with the implicit comparison group being monolinguals who use one language variety in all situations with all interlocutors. The present study, therefore, compares heritage speakers’ language use, exposure, ability, and identity to other HL-ML bilingual adults living and working in an HL dominant society (late bilinguals in the US) to appropriately contextualize the heritage speakers’ sociolinguistic profile. The present study will compare US Spanish heritage speakers born in the anglophone US or brought to the anglophone US before the age of 8 to late bilinguals who emigrated to the anglophone US after age 18. These late bilinguals function as the apparent-time parents of the heritage speakers in the study. A principal component analysis (PCA) of the sociolinguistic questionnaire responses is run and interpreted. The PCA provides a profile of US Latinx Spanish-English bilinguals in showing how features of the bilinguals’ early language exposure patterns with current language choice in different contexts and with different interlocuters in distinct ways. The benefit of the PCA-derived variables is that these interrelated behaviors are separated into distinct unrelated variables. These derived variables of language use, exposure, ability, and identity are then compared across group
to identify dimensions on which heritage speakers differ from and are similar to late bilinguals in their sociolinguistic profile.

3.1 Methodology

3.1.1 Participants

Eighty-two Spanish-English bilingual adults living in the New York City participated in the study at the CUNY Graduate Center in midtown Manhattan, NY and were compensated financially for their participation. The study protocol was approved by the CUNY institutional review board and written consent was obtained.

Native Spanish-speaking participants were categorized as either heritage Spanish speakers (n=43) or late bilinguals (n=39) based on criteria commonly used in heritage speaker studies (Benmamoun, Montrul, & Polinsky, 2013) and pre-determined inclusion criteria. Nearly three-quarters of the heritage speakers were born in the Anglophone US (65.12%, n=28) and the rest moved to the Anglophone US before age 8 (34.88%, n=15, M=3.31, SD=2.29). Heritage speakers were raised speaking primarily Spanish until at least age 10 by Spanish-speaking immigrant parents originally from a Spanish-dominant country/region. Late bilinguals were born in a Spanish-dominant country/region and moved to the Anglophone US at the age of 17 or older (M=25.74, SD=4.70).

Both groups were adults younger than 45 years old (range: [18, 44]), although heritage speakers (M=25.21, SD=6.41) were significantly younger than late bilinguals (M=30.31, SD=5.33), t(80)=3.90, p<.001. There was also a significant group difference in education. Education was quantified as numericized highest degree completed: 0=HS, 5=PhD. Late bilinguals (M=3.61, SD=0.93) were significantly more educated than heritage speakers (M=2.58, SD=1.07), t(72)=4.31, p<.001. More late bilinguals (n=22; 66.67%) had masters or doctorate degrees than heritage speakers (n=10; 24.39%). Two participants were excluded from the analysis due to having too much missing data in the survey.

3.1.2 Protocol

The language background questionnaire was administered during the recruitment processes for the current studies in the Second Language Acquisition Lab at the CUNY Graduate Center. The language
background questionnaire consists of 68 questions that the research assistants asked in English over the phone and in-person. The questionnaire included items probing language history (Li, Sepanski, & Zhao, 2006) and additional items probing demographics (education level, social class), language ability (self-reported literacy and speaking abilities in English and Spanish), current language use (language used with different interlocutors and contexts), and language exposure (language used during childhood and adolescence, amount of exposure to English over the lifetime). Responses were input and checked by at least two RAs and coded in R (R core team, 2018).

3.1.3 Variables of interest

The present study examines responses from 26 of the 68-item questionnaire. Items are listed in appendix A. Four questions probed language ability: Spanish literacy, Spanish speaking ability, English literacy, English speaking ability. Participants were asked to rate their language ability on a 5-point Likert scale with the endpoints labeled 1=limited knowledge to 5=native. The fourth point on the scale was labeled 4=fluent. Ability variables were re-coded so that values ranged from 1 to 4 and responses that were 4=fluent and 5=native were collapsed as both indicate fluent use of the language. Two more dichotomous variables were derived from the speaking ability variables where 1=native and 0=non-native which reflects self-identification as a native speaker of Spanish and/or English.

Eight variables probed language used with different interlocutors: father, mother, partner, sibling, boss, classmate, coworker, friends, younger children. Two variables probed language spoken in the home growing up and which languages were used the most often. Responses were numericized as 0=Spanish, 0.5=Spanish and English, 1=Spanish. The variable probing language used with younger children was excluded due to insufficient data. For individuals responding, late bilinguals (n=4) exclusively used Spanish and heritage speakers (n=2) reported using a mix of Spanish and English or English only with younger children. Three variables probed language used when reading, listening to music/radio, and watching TV and three variables probed language used in different contexts: home, social environments, work. Responses were numericized as 0=no Spanish, 0.5=English and Spanish, 1=mostly Spanish.

The variable that indexes the primary language spoken in the local community from ages 5-18 and the variable that indexes the language of learning and teaching from ages 5-18 were quantified as
the mean numericized language reported by year with 0=Spanish, 0.5=Spanish and English, 1=Spanish.

The variable of cumulative exposure to English was calculated as the proportion of the participant's life in an English dominant environment (age of arrival in Anglophone US for late bilinguals or age of entering English LoLT school system for heritage speakers).

3.2 Principal component analysis

To explore the relationship between the items in the questionnaire, a principal component analysis was run on the correlation matrix. Correlations between the 25 variables are in Figure 1. The strength of the correlations are indicated in red. The PCA uses the correlations between the variables to group variables into independent constructs that will used to look at group differences. The PCA was performed for responses from 25 variables of language use and exposure across bilingual group with the psych::principal() function in R (Revelle, 2016). The stopping criterion to determine the number of rotated components that should be retained for the solution with the varimax procedure was the Kaiser-Guttman rule (Cangelosi & Goriely, 2007). The varimax rotation was used so that the retained rotated components were uncorrelated or orthogonal.

The 8 component PCA solution on 25 variables was examined and items that had primary component loadings less than |.50| (n=3; language spoken in youth, language currently used at home, language currently used in social settings) and items that had cross-loadings greater than |.50| (n=1; language used in youth) were excluded. The PCA was re-run retaining 7 rotated components and reported below.
### 3.2.1 Rotated components

In the PCA final solution, seven rotated components were retained, cumulatively accounting for 66% of the total variance in the data set. All items had a primary loading over |.50|. This rotation created seven factors with sums of squared loadings ranging from 1.35 to 4.00 and the mean item complexity was 2.20. Communalities of the items ranged from .49-.87. The root mean square of the residuals is 0.06 (empirical $\chi^2=186.72, \ p<.05$). The component loadings greater than |.20| can be seen in Table 1 and visualized in Figure 2.

![Figure 1: Correlation matrix of language use, behavior, and exposure variables across group. Strength of correlation indicated with red shading.](image-url)
Figure 2: Factor loadings by language background questionnaire items by component. Red=negative values, blue=positive values. Strength of factor loading represented by transparency.

The first rotated component accounted for 16% of variance and the variables that loaded highly on the component indexed language used at work, eigenvalue=4.00. The variables that loaded the highest were language used with the participants’ boss, classmate, coworker, and at work. The second rotated component accounted for 14% of variance and the variables that loaded highly on the component indexed native English speaker identity, eigenvalue=3.39. The variables that loaded the highest were identifying as a native English speaker, language used with siblings, cumulative English used, language of community in youth, and language used when reading. The third rotated component accounted for 10% of variance and the variables that loaded highly on the component indexed native Spanish speaker identity, eigenvalue=2.47. The variables that loaded the highest were identifying as a native Spanish speaker, language used with parents, and Spanish literacy ability. Higher Spanish literacy ability and self-identification as a native Spanish speaker is associated with speaking Spanish with their father and mother and speaking English with a partner.
<table>
<thead>
<tr>
<th>Language exposure</th>
<th>Work</th>
<th>English Identity</th>
<th>Spanish Identity</th>
<th>English Ability</th>
<th>Media</th>
<th>English Use</th>
<th>Spanish Ability</th>
<th>Commun-alities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative English Identity</td>
<td>0.68</td>
<td>0.44</td>
<td>0.22</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td>0.82</td>
</tr>
<tr>
<td>English in Community</td>
<td>-0.47</td>
<td>-0.24</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.63</td>
</tr>
<tr>
<td>English as LoLT</td>
<td>-0.24</td>
<td></td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.73</td>
</tr>
<tr>
<td>Current language use</td>
<td>Father</td>
<td>0.29</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>Mother</td>
<td></td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Partner</td>
<td>0.31</td>
<td>0.40</td>
<td>-0.23</td>
<td>0.50</td>
<td>-0.21</td>
<td>0.44</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Sibling</td>
<td></td>
<td>0.84</td>
<td>0.27</td>
<td>0.23</td>
<td></td>
<td></td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Boss</td>
<td>0.75</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>Classmate</td>
<td>0.67</td>
<td>0.35</td>
<td></td>
<td>-0.25</td>
<td></td>
<td></td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>Coworker</td>
<td></td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>Friends</td>
<td>0.36</td>
<td>0.41</td>
<td>0.21</td>
<td>-0.28</td>
<td>0.33</td>
<td></td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Reading</td>
<td>-0.38</td>
<td>-0.57</td>
<td></td>
<td>0.26</td>
<td></td>
<td>0.27</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>Radio &amp; Music</td>
<td></td>
<td></td>
<td></td>
<td>0.82</td>
<td></td>
<td></td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>TV</td>
<td></td>
<td></td>
<td></td>
<td>0.83</td>
<td></td>
<td></td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>Work</td>
<td>-0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language ability</td>
<td>English Native Ident</td>
<td>0.77</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>English Speaking</td>
<td></td>
<td></td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>English Literacy</td>
<td>0.22</td>
<td></td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>Spanish Native Ident</td>
<td>-0.34</td>
<td>-0.48</td>
<td></td>
<td>-0.30</td>
<td></td>
<td></td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Spanish Speaking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.91</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Spanish Literacy</td>
<td></td>
<td></td>
<td>-0.74</td>
<td></td>
<td>0.32</td>
<td></td>
<td>0.70</td>
</tr>
</tbody>
</table>

Table 1: Factor loading by language background questionnaire items by component greater than |.40|.

The fourth rotated component accounted for 8% of variance and the variables that loaded highly on the component indexed English ability, eigenvalue=1.98. The variables that loaded the highest were English speaking ability and English literacy ability. Higher English speaking and literacy abilities were associated with English used with a partner. The fifth rotated component accounted for 7% of variance and the variables that loaded highly on the component indexed language used consuming media,
eigenvalue=1.73. The variables that loaded the highest were language used when watching TV and listening to music or the radio. The sixth rotated component accounted for 6% of variance and the variables that loaded highly on the component indexed early English use, eigenvalue=1.52. The variables that loaded the highest were language used in school, language used in the community, language used with partner, and language used with friends. The more English that NYC Spanish-English bilinguals were exposed to in school as the LoLT and in the community, the more English they currently use with their partner and friends. The seventh rotated component accounted for 5% of variance and the variable that loaded highly on the component indexed Spanish ability, eigenvalue=1.35. The variable that loaded the highest was self-rated Spanish speaking ability.

3.2.2 Group comparisons

To explore the similarities and differences in language use, exposure, and ability between late bilinguals and heritage speakers, group differences for the seven rotated components scores are compared with independent samples t-tests in R. If equal variances cannot be assumed due to a significant variance test, the Welch’s t-test is run.

![Use Variables by Group](image)

**Figure 3:** Mean component scores for work, media, and English use factors by bilingual group with 95% confidence interval error bars.
The mean scores for the three components that indexed primarily language use by- and across-group are in Table 2 and plotted in Figure 3. Heritage speakers used more English at work and when consuming media than late bilinguals. Heritage speakers \( (M=0.22, SD=0.66) \) had significantly higher work English scores than late bilinguals \( (M=-0.24, SD=1.21) \), \( t(56.01)=-2.07, p<.05 \). Heritage speakers used more English with their boss, coworkers, classmates and at work. Heritage speakers \( (M=0.22, SD=0.98) \) had significantly higher media English scores than late bilinguals \( (M=-0.25, SD=0.97) \), \( t(78)=-2.16, p<.05 \). Heritage speakers used more English when watching TV and listening to music or the radio. Heritage speakers \( (M=0.03, SD=0.90) \) did not significantly differ from late bilinguals \( (M=-0.04, SD=1.11) \) in terms of general English use scores \( (t(78)=-0.31, p=.76) \) as both groups had scores near zero. However, differences were in the expected direction: heritage speakers’ general English use scores were slightly positive and late bilinguals’ general English use scores were slightly negative.

<table>
<thead>
<tr>
<th></th>
<th>Late bilinguals</th>
<th>Heritage Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work</td>
<td>-0.24</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td>(0.66)</td>
</tr>
<tr>
<td>Media</td>
<td>-0.25</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(0.97)</td>
<td>(0.98)</td>
</tr>
<tr>
<td>English Use</td>
<td>-0.04</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(1.11)</td>
<td>(0.90)</td>
</tr>
</tbody>
</table>

**Table 2:** Mean component scores for work, media, and English use factors by bilingual group. Significance of t-test for group means. Standard deviations in parentheses.

The mean scores for the three components that primarily indexed native speaker identity by- and across-group are in Table 3 and plotted in Figure 4. Heritage speakers \( (M=0.67, SD=0.70) \) had significantly higher English identity scores than late bilinguals \( (M=-0.74, SD=0.68) \), \( t(78)=-9.07, p<.001 \). Heritage speakers identified more as native English speakers, spoke more English with their siblings, read in English more, and were exposed to more English in their community and as a proportion of their life. Heritage speakers \( (M=-0.35, SD=1.21) \) had significantly lower Spanish identity scores than late bilinguals \( (M=0.39, SD=0.48) \), \( t(54.55)=-3.70, p<.001 \). Heritage speakers identified less as native Spanish speakers, self-rated lower on Spanish literacy skills, and spoke more English with their father, and mother.
**Figure 4:** Mean component scores for English and Spanish identity factors by bilingual group with 95% confidence interval error bars.

<table>
<thead>
<tr>
<th>Component</th>
<th>Late Bilinguals</th>
<th>Heritage Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Identity</td>
<td>-0.74</td>
<td>0.67 ***</td>
</tr>
<tr>
<td></td>
<td>(0.68)</td>
<td>(0.70)</td>
</tr>
<tr>
<td>Spanish Identity</td>
<td>0.39</td>
<td>-0.35 ***</td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
<td>(1.21)</td>
</tr>
</tbody>
</table>

**Table 3:** Mean component scores for English and Spanish identity factors by bilingual group. Significance of t-test for group means. Standard deviations in parentheses.

The mean scores for the three components that indexed primarily language ability by- and across- group are in Table 4 and plotted in Figure 5. Heritage speakers and late bilinguals had similar English and Spanish ability, although there was more within-group variation in the non-dominant language. Heritage speakers ($M=0.14$, $SD=0.24$) and late bilinguals ($M=-0.16$, $SD=1.42$) did not have significantly different English ability scores, $t(38.93)=1.27$, $p=.21$. There was more variability in late bilinguals' English ability scores. The non-significance of group differences for English ability could be related to the study sample; late bilinguals were older and more educated than heritage speakers (see section 1.4.1.1). This might bias the results, as the prototypical parents of heritage speakers are not as well educated. Heritage speakers ($M=-0.03$, $SD=1.25$) and late bilinguals ($M=0.03$, $SD=0.66$) also did not have significantly
different Spanish ability scores, $t(63.32)=-0.25$, $p=.80$. There was more variability in heritage speakers' Spanish ability scores.

![Figure 5: Mean component scores for English and Spanish ability factors by bilingual group with 95% confidence interval error bars.](image)

**Table 4**: Mean component scores for English and Spanish ability factors by bilingual group. Significance of t-test for group means. Standard deviations in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Late Bilinguals</th>
<th>Heritage Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English ability</strong></td>
<td>-0.16 (1.42)</td>
<td>0.14 (0.24)</td>
</tr>
<tr>
<td><strong>Spanish ability</strong></td>
<td>0.03 (0.66)</td>
<td>-0.03 (1.25)</td>
</tr>
</tbody>
</table>

3.3 Discussion

The PCA provides an empirical profile of US Latinx Spanish-English bilinguals in showing how features of the bilinguals’ early language exposure pattern with current language choice in different contexts and with different interlocuters in distinct ways. The benefit of the PCA-derived variables is that these interrelated behaviors are separated into distinct constructs. Heritage speakers and late bilinguals differed in language use, exposure, and identity for some of these distinct constructs, but were similar in language ability and language use associated with ability. Heritage speakers, who were raised in the anglophone
US.

Each group also differed in their identification with their dominant language: English for heritage speakers, Spanish for late bilinguals. Identity was also associated with the language Spanish-English bilinguals used with their family. Identification as a Spanish native speaker was associated with using more Spanish with one’s parents and identification as an English native speaker was associated with using more English with one’s siblings. Both types of bilinguals are fluent users of English and Spanish. They were similar in their self-rated abilities in English (speaking and literacy) and Spanish (speaking only) and the language they used with their friends and partner. It is not surprising that Spanish literacy ability did not pattern with Spanish speaking ability as heritage speakers were educated in English-medium schools, and that higher self-rated Spanish speaking ability was associated with more use of Spanish with friends and partners. Heritage speakers are typically characterized as being dominant in the social majority language whereas late bilinguals are typically characterized as being dominant in the social minority language. These findings suggest that “dominance” is related more to identity than actual ability, as both groups had similar self-rated abilities in both languages. Lastly, the Latinx anglophone US-based bilingual groups did not differ from each other in general English use. This finding is not surprising given that all individuals live in the anglophone US, which requires regular use of the English language in order to work, shop, commute, and generally live life. Even if individuals live and work in communities with dense and large co-ethnic populations, one still must use the social majority language regularly. This analysis demonstrates that while heritage speakers differ from their apparent-time parents in self-identification and language use, they did not differ in their command of both languages.

4. Present issues with the studies of heritage speakers

The dearth of studies of heritage speakers’ linguistic knowledge is a critical issue as they are currently facing many challenges (principally in education) due to an inadequate understanding of their abilities and the use of inappropriate comparison group. Heritage speakers of Spanish in the anglophone US find themselves facing standards and expectations based on prescriptive and literary standards and norms that they would not have been exposed to during colloquial conversations. Additionally, many of these
speakers live in areas where undervalued varieties of English are spoken, and their English skills are thus
different from the prescriptive norm. However, as children who have parents who speak a language other
than English in the home are required to test out of the English Language Learner designation, this has
led to the disheartening characterization by some educators that these heritage speakers are not
proficient in Spanish, and also not proficient in English. With increased teacher education and advocacy,
the attitude toward this population is slowly beginning to shift away from this deficit-based perspective.
But still these inappropriate comparisons of heritage speaker grammar to prestige varieties to which
heritage speakers are not exposed are hindered rather than helped by academic assertions of incomplete
acquisition rather than differential acquisition. Legislation (IDEA, 2004) now requires heritage language
skills to be included in language assessments, and forbids students being place in special education
programs based on limited exposure to or proficiency in English or speaking a variety of English or the
heritage language that is not a prestige variety. Professional organizations (American Speech-Language-
Hearing Association, 2017) and state-level legislation such as the New York State Education Department
(2010) have followed suit in their policy recommendations However, these policies have not yet made
their way into practice in the classroom contexts, and many educators continue to operate under the
outdated understanding of heritage speakers as lacking proficiency in both of their languages.

Inquiries into the linguistic knowledge of heritage speakers are relatively new; with early work
appearing in the early 2000s (for example, the Heritage Language Journal began publication in 2002).
The term “heritage language” was first used in the Canada in 1977 (Kagan & Dillon, 2008) in the context
of First Nation language programs. As much of the research about heritage language speakers is
conducted from an education perspective, relatively few studies have been concerned with the theoretical
aspects of heritage language, and investigations into morphosyntactic processing have only recently
begun (see Zhang (2017) for a review). Benmamoun, Montrul, & Polinsky (2010) characterize heritage
speakers as language users with limited HL competence, whose “inflectional morphology and complex
syntax are highly vulnerable to attrition and incomplete acquisition” (p. 13). Incomplete acquisition is
defined as the failure to reach adult-like levels of grammatical competence and is attributed to insufficient
heritage language input prior to age twelve. Attrition is defined as a loss of linguistic knowledge that has
already been acquired, and is attributed to reduced frequency of heritage language use. Montrul (2005)
characterizes the heritage speaker’s grammar as “incomplete or eroded” and as “typical of intermediate or advanced L2 learners”. Scontras, Fuchs, and Polinsky (2015) characterize the heritage speaker’s grammar as “simpler than the native baseline” and as having “reduced complexity” and “reduced expressive power” when compared to a native baseline. Benmamoun, Montrul, and Polinsky (2010) acknowledge the heterogeneity among heritage speakers, but frame the debate around whether the differences between heritage speakers and monolinguals are due to incomplete acquisition, or attrition, or both, which leads to further questions of which structures are lost through attrition and which structures are not acquired at all.

4.1 Prescriptivist and monolingual comparison

One of the prominent early works on heritage speakers is Valdés (1989), which is an examination of the Spanish skills of Spanish heritage language speakers in the US. This work raises the issue that heritage language speakers are often classified as low-proficiency in Spanish, but that this finding is usually associated with a comparison to an educated monolingual Spanish speaker who speaks a prestige variety. Valdés calls into question the concept of a “native speaker” and argues that Spanish heritage speakers in the US, who may speak undervalued varieties of Spanish, are held to inappropriate standards, and therefore the notion that heritage speakers are low-proficiency speakers is also inappropriate. Valdés calls for heritage language assessment to become compatible with the varieties that are actually spoken by heritage speakers and compatible with the oral modality and non-academic context that will give a true indication of heritage language skills.

Despite this early work advocating comparing heritage language skills to the ambient language variety in an appropriate context and modality, the native speaker monolingual standard continues to be used as the point of comparison for heritage speakers’ language skills in most linguistic research. Scontras, Fuchs, and Polinsky (2015) even refer to the “establishment of a clear native baseline” as “a must for any comparison” (p. 16) which 1) makes the erroneous assumption that heritage speakers are not native speakers of their HL, and 2) assumes that the base of comparison is a monolingual without addressing issues with prescriptivism, literacy, and language variety that makes the use of such a reference group inappropriate. Monolinguals are an inappropriate comparison group for a number of reasons, including “Given coactivation and bidirectional effects, neither the first nor the second language
of bilinguals can be expected to resemble under scrutiny that of monolinguals in either language” (Birdsong, 2018, p. 6). Additionally, the native monolingual standard is likely not the language variety that heritage speakers are exposed to in either of their languages. Educated native speakers learn prescriptive grammatical structures for written language in school that has a standardizing effect of language variation. Flores (2014) argues that heritage speakers are not unique in being principally exposed to informal or colloquial registers. Monolinguals who have similarly limited contact with prestige-variety-medium education are less accurate on properties of the formal register. However, many people continue to believe that these prescriptive structures, features of the formal registers, that are uncommon or absent in speech are critical components to the “complete” acquisition of a language. A comparison of heritage speakers of Brazilian Portuguese and monolingual speakers (Rothman, 2007) showed that heritage speakers who regularly used Brazilian Portuguese in daily life had a lower performance on a test of inflected infinitives than heritage speakers who were formally taught Brazilian Portuguese. This had previously been characterized as a failure to acquire monolingual-like knowledge of inflection. However, Rothman explains that speakers who used Brazilian Portuguese but were not educated in the formal written variety should not be expected to have that structure in their speech. A follow-up study (Rothman, 2009) additionally showed that heritage speakers who regularly used European Portuguese in daily life but who were not exposed to the formal written variety used inflected infinitives similarly to monolinguals. Therefore, when undertaking a comparison to monolinguals, language variety and influence of written structures must be taken into account, as these variables may account for the differences which have been held as examples of incomplete competence.

4.2 Linguistic insecurity

In addition to issues with prescriptivist and monolingual comparisons, the validity of using metalinguistic judgment tasks with heritage speaker populations can be held suspect because of the sense of (meta-)linguistic insecurity that heritage speakers experience in the heritage language. Linguistic insecurity that manifests as poor performance on metalinguistic tasks is due to the explicit and implicit bias against bilingualism in the United States.
Heritage speakers anecdotally describe feeling “non-native” in both of their languages. This comes from teacher feedback at school that their English skills are behind and they need to catch up, or from failing standardized tests of English, despite speaking English more frequently than their home language as they progress in their schooling. In New York City schools, the long-term English language learner population is likely composed of such students. Additionally, within the community, heritage language speakers may be at the same time encouraged by parents to focus on mastering the societally dominant language, yet simultaneously shamed by their ethnolinguial community for losing their mother tongue. Ridicule and shaming of less proficient HL speakers, further increases linguistic insecurity and discourages HL use (Cho, Shin, & Krashen, 2004).

There has been little overt experimental linguistics research on linguistic insecurity, and much of it is based on anecdotal evidence (Klein & Martohardjono 2006) or on post-hoc conjectures about different patterns of performance on rating tasks (e.g., Polinsky, 2006; Scontras, Fuchs, & Polinsky, 2015). Klein and Martohardjono (2008; 2009) found that adolescent Spanish speakers with low literacy levels in their native Spanish who recently arrived in the United States hesitated during metalinguistic judgment tasks, even when oral proficiency and fluency were high. Furthermore, Benmamoun, Montrul, & Polinsky (2010) report that heritage speakers tend to misjudge their own ability in the home language. Fluent speakers may self-report lower proficiency, and speakers with less proficiency may report better language skills than they actually have.

Heritage speakers’ patterns of responses to metalinguistic tasks have been characterized as exhibiting a “yes-bias” in that they accept grammatical structures and ungrammatical structures (Polinsky, 2016 discussing results from Polinsky, 2006). Polinsky (2016) conjectures that heritage speakers are “aware of their limitations in their knowledge (constantly being reminded how little they know…) and are therefore unprepared to reject unfamiliar grammatical structures, assuming they are observing a grammatical form they have simply not encountered yet” (p. 12). Evidence of this “yes-bias” has been argued to come from grammatical/ungrammatical sentence ratings in studies of heritage speakers of Russian (Polinsky, 2006, 2016), Mandarin (Scontras, Badecker, Shank, Lim, & Fedorenko, 2015), Japanese and Korean (Laleko & Polinsky, 2013, 2016) in the anglophone US; Labrador Inuittitut in anglophone Canada (Sherkina-Lieber, 2011); and English in hebraiophone Israel (Vishwanath, 2013).
Additional evidence of the possible “yes-bias” comes from Heidrick (2017) who reported good/terrible ratings of prescriptively documented collocations and new collocations by Spanish heritage speakers in the anglophone US. They found that Spanish heritage speakers rated new collocations as better than Spanish monolinguals, and Spanish-dominant Spanish-English bilinguals in Spanish-dominant societies and English-dominant societies.

In addition to the "yes-bias" argued for above as a post-hoc rationalization of heritage speaker differences in rating from monolingual controls, and the hesitation in metalinguistic ratings from low-literacy L1-dominant bilinguals (which is informative given the low-formal Spanish literacy skills of heritage speakers), heritage speakers are also reported to qualitatively respond “maybe” or “I don’t know” when asked about the acceptability of ungrammatical sentences (Polinsky, 2016 discussing results from Polinsky, 2006). However, avoidance of the edges of a Likert scale item is an idiosyncratic response to a Likert scale item. Some individuals display “fence-sitting” behaviors when asked for Likert scale item ratings in a number of disciplines such as education, and in the medical field (Brown, 2000). The edge-avoidance of heritage speakers should therefore not be taken to reflect insecurity of their HL linguistic knowledge, but instead is a common confound of the methodology itself.

The confluence of the heritage speaker "yes-bias", Likert scale “maybe-bias”, and low-literacy hesitancy make metalinguistic tasks wholly inappropriate for probing the grammar or language processing of heritage speakers. Additionally, the studies above almost exclusively use Likert-type items for the ratings and then analyze these results of the studies treating the data as continuous interval data with statistical methods that assume a Gaussian (normal) distribution. Whether Likert-type item ratings should be treated as continuous interval or ratio data with parametric tests, or categorical data with non-parametric tests is outside the scope of this study (see Brown (2011) for an overview of the debate). As is common practice in experimental syntactosemantic studies in generative linguistics, we will consider the Likert-type item rating data to be continuous interval data. Likewise, the exact relationship between grammaticality and acceptability (Sprouse, 2007a; Myers, 2017), the debate about whether grammaticality can and/or should be measured as continuous and gradient (Sprouse, 2007b; Sorace & Keller, 2005), whether Likert type item ratings are appropriate (Sprouse & Almeida, 2017; Langsford, Perfors, Hendrickson, Kennedy, & Navarro, 2018), or whether grammaticality is fundamentally categorical
and only surfaces as gradient due to measurement protocols (Coetzee, 2009; Gorman, 2014) is outside the scope of the present study. At issue, is whether one ought to use parametric tests when looking for group differences on Likert-type item data. There are two major issues with the data: 1) it is highly skewed (Jamieson, 2004), 2) it is highly polarized (ibid.) in that the majority of responses are often at the end points of the scale, and 3) it is bounded which creates floor/ceiling effects (Clason & Dormody, 1994). Common statistical tests used in linguistic, such as t-tests, ANOVAs, (multiple) linear regressions, and linear mixed-effects models, assume a Gaussian (normal) distribution that is not bounded. This can lead to values of the estimates of ratings outside of the limits of the measurement scale when using regression modeling (e.g., a predicted 0.5 rating or 8.0 rating when the measurement scale used included only values 1-7). While Clason & Dormody (1994) argue that individual item ratings cannot be analyzed and instead the items must be summed, the present study examined the aggregate ratings from Likert-type items. The distribution of the item ratings most closely approximates a non-Gaussian distribution: the beta distribution.

4.3 Empirical study of linguistic insecurity

In order to explore the effects of heritage speaker HL linguistic insecurity using appropriate statistical techniques, Likert-type item naturalness ratings of grammatical and ungrammatical noun phrase wh-extraction from different types of islands in Spanish from Spanish heritage speakers and Spanish L1, English L2 late learners (late bilinguals) are modeled with a zero-and-one inflated beta regression (Ferrari & Cribari-Neto, 2004). If a heritage speaker “yes-bias” is found, we would see heritage speakers accepting both grammatical and ungrammatical items as natural. The late bilinguals, as a comparison group, are expected to rate grammatical items as natural and ungrammatical items as unnatural given their L1 dominance, education in the L1, and late exposure to the L2.

4.3.1 Methodology

4.3.1.1 Participants

Seventy-seven Spanish-English bilingual adults living in the New York City participated in the study at the CUNY Graduate Center in midtown Manhattan, NY and were compensated financially for their
participation. The CUNY institutional review board approved the study protocol, part of a larger study of syntactic processing in NYC Spanish-speaking Latinxs, and written consent was obtained.

Native Spanish-speaking participants were categorized as either heritage Spanish speakers \((n=41)\) or late bilinguals \((n=36)\) based on criteria commonly used in heritage speaker studies (Benmamoun, Montrul, & Polinsky, 2013) and the pre-determined inclusion criteria. Over two-thirds of the heritage speakers were born in the Anglophone US \((68.29\%, n=28)\) and the rest moved to the Anglophone US before age 8 \((31.71\%, n=13, M=3.43, SD=2.37)\). Heritage speakers were raised speaking primarily Spanish until at least age 10 by Spanish-speaking immigrant parents originally from a Spanish-dominant country/region. Late bilinguals were born in a Spanish-dominant country/region and moved to the Anglophone US at the age of 17 or older \((M=25.91, SD=4.82)\).

### 4.3.1.2 Stimuli

The auditory stimuli were Spanish complex sentence dyads: a declarative context sentence, and a related wh-question. The target wh-question was either prescriptively grammatical or ungrammatical. There were five items per condition per grammaticality (except for TTSE with 7 items) for the 10 conditions (total items \(n=104\)): complex noun phrase complement islands with object wh-extraction (CCOE) in (1), complex noun phrase complement islands with subject wh-extraction (CCSE) in (2), object-object relative clause island with subject wh-extraction (RCOO) in (3), object-subject relative clause island with object wh-extraction (RCOS) in (4), object-subject relative clause island with object wh-extraction (RCSS) in (5), temporal adverbial island with object wh-extraction (TAOE) in (6), temporal adverbial island with subject wh-extraction (TASE) in (7), complement clause island with subject wh-extraction (TTSE) in (8), wh-island with object wh-extraction (WHOE) in (9), and wh-island with subject wh-extraction (WHSE) in (10).

**Complex noun phrase complement islands with object wh-extraction (CCOE)**

(1) a. El gobierno afirmó la creencia que el ejército produce un arma biológica.  
*The government affirmed the belief that the army produced a biological weapon.*

b. ¿Qué arma biológica afirmó el gobierno que el ejército produce?  
*What biological weapon did the government affirm that the army produced?*

c. *¿Qué arma biológica afirmó el gobierno la creencia que el ejército produce?*  
*What biological weapon did the government affirm the belief that the army produced?*
Complex noun phrase complement islands with subject wh-extraction (CCSE)
(2) a. Juan contó el chisme que el vecino robó el carro anoche.
\[Juan told the gossip that the neighbor stole the car last night\]
   b. ¿Qué vecino contó Juan que robó el carro anoche?
   \[What neighbor did Juan tell that stole the car last night?\]
   c. *¿Qué vecino contó Juan el chisme que robó el carro anoche?
   \[*What neighbor did Juan tell the gossip that stole the car last night?*

Object-object relative clause island with subject wh-extraction (RCOO)
(3) a. La ciudad clausuró la playa que el huracán inundó.
\[The city closed the beach that the hurricane flooded\]
   b. ¿Qué ciudad clausuró la playa que el huracán inundó?
   \[What city closed the beach that the hurricane flooded?\]
   c. *¿Qué huracán clausuró la ciudad la playa que inundó?
   \[*What hurricane did the city close the beach that flooded?*

Object-subject relative clause island with object wh-extraction (RCOS)
(4) a. Paola hizo el gesto que causó la controversia.
\[Paola made the joke that caused the controversy\]
   b. ¿Qué gesto hizo Paola que causó la controversia?
   \[What joke did Paola make that caused the controversy?\]
   c. *¿Qué controversia hizo Paola el gesto que causó?
   \[*What controversy did Paola make the joke that caused?*

Subject relative clause island with object wh-extraction (RCSS)
(5) a. La investigadora que descubrió la molécula explicó su método.
\[The researcher that discovered the molecule explained her methodology\]
   b. ¿Qué método explicó la investigadora que descubrió la molécula?
   \[What methodology did the researcher explain that discovered the molecule?\]
   c. *¿Qué molécula la investigadora que descubrió explicó su método?
   \[*What molecule did the researcher that discovered explain her methodology?*

Temporal adverbial island with object wh-extraction (TAOE)
(6) a. El conductor paró el carro mientras que el niño cruzaba la calle.
\[The driver stopped the car while the child crossed the street\]
   b. ¿Qué carro paró el conductor mientras que el niño cruzaba la calle?
   \[What car did the driver stop while the child crossed the street?\]
   c. *¿Qué calle paró el conductor el carro mientras que el niño cruzaba?
   \[*What street did the driver stop the car while the child crossed?*

Temporal adverbial island with subject wh-extraction (TASE)
(7) a. El niño comió el dulce mientras que su tía buscaba la comida.
\[The child ate the candy while his aunt looked for food\]
   b. ¿Qué niño comió el dulce mientras que su tía buscaba la comida?
   \[What child ate the candy while his aunt looked for food?\]
   c. *¿Qué tía el niño comió el dulce mientras que buscaba la comida?
   \[*What aunt did the child eat the candy while looked for food?*

Complement clause island with subject wh-extraction (TTSE)
(8) a. El buzo exclamaba que un tiburón había mordido su tanque de oxígeno.
\[The diver shouted that the shark had bitten his oxygen tank\]
   b. ¿Qué tiburón exclamaba el buzo que había mordido su tanque de oxígeno?
   \[What shark did the diver shout that had bitten his oxygen tank?\]
   c. ¿Qué tiburón exclamaba el buzo había mordido su tanque de oxígeno?
   \[What shark did the diver shout had bitten his oxygen tank?\]
Wh-island with object wh-extraction (WHOE)
(9) a. Ignacio confirmó por qué la enfermera había llevado la medicina.
   Ignacio confirmed why the nurse had brought the medicine.
   *Ignacio confirmed why the nurse had brought the medicine.
   b. ¿Qué medicina confirmó Ignacio que la enfermera había llevado?
   What medicine did Ignacio confirm that the nurse had brought?
   *What medicine did Ignacio confirm that the nurse had brought?
   c. *¿Qué medicina confirmó Ignacio por qué la enfermera había llevado?
   *What medicine did Ignacio confirm why the nurse had brought?

Wh-island with subject wh-extraction (WHSE)
(10) a. Juan odia cómo el profesor administra el examen.
   Juan hated how the professor administered the exam
   *Juan hated how the professor administered the exam
   b. ¿Qué profesor odia Juan que administra el examen?
   What professor did Juan hate that administered the exam?
   *What professor did Juan hate that administered the exam?
   c. *¿Qué profesor odia Juan cómo administra el examen?
   *What professor did Juan hate how administered the exam?

Stimuli were recorded in a sound proof booth by a female L1 Spanish late bilingual speaker from Mexico using SoundForge as natural running speech with neutral prosody and sampled at 44.1kHz. All stimuli were normalized, amplified to an average loudness of -26.00 dB, and the noise filtered out using Audacity® 2.0.3 (Audacity Team, 2014), then exported as WAV files.

4.3.1.3 Protocol
The acceptability judgment task was administered 10-14 days after the ERP or pupillometry experiment, the primary task that participants were recruited for. During this task, participants rated the acceptability of a subset of the trials from the ERP or pupillometry experiments. The stimuli were presented aurally in sentence dyads: a declarative context sentence followed by either an ungrammatical wh-question or a grammatical wh-question. Following presentation of the dyads, participants rated the naturalness of the wh-question (the second sentence) using a five-point Likert scale with the ends and midpoints labeled in Spanish (1 = es natural “is natural”, 3 = puede ser natural “possibly natural”, and 5 = no es natural “is unnatural”). Participants were instructed to respond to the sentences as quickly as possible by pressing a button on a serial response box. There was a 1000ms inter-stimulus interval. Items were pseudorandomized over two blocks, each lasting approximately 20 minutes with a short break in between. Trials were presented over external speakers using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) in a quiet room in the SLA lab following a practice session. Instructions were presented in both Spanish and English.
4.3.1.4 Data analysis

Participant-average rating was plotted and calculated in R to explore trends in the data. Rating data are bounded between values of 1 and 5 and had a non-Gaussian distribution. The bounded asymmetrical rating distribution is most similar to a beta-distribution, common with rates and proportion. When regression models assuming a Gaussian distribution are fit to data with a beta distribution, the results are heteroskedastic where there is more variation around the mean and less variation at the upper and lower limits (Cribari-Neto & Zeileis, 2004). A regression model appropriate for beta-distributed data provides estimates for the mean $\mu$ and the dispersion $\sigma$ or precision $\varphi$ parameters (Ferrari & Cribari-Neto, 2004).

The beta regression model has been extended to include responses at the limits of the scale, not just values between the end points, referred to as an inflated beta regression, zero-inflated if the data include values of the lower bound, one-inflated if the data include values of the upper bound, and zero-and-one inflated if the data include responses of the upper and lower bound (Ospina & Ferrari, 2010, 2012). The naturalness rating data are bounded $1 \leq y \leq 5$. This was transformed to values $0 \leq y \leq 1$ and fit with a zero-and-one inflated beta regression using the \texttt{gamI:gamI:amss()} function (Rigby & Stasinopoulos, 2005).

Formula for the four distribution parameters ($\mu=0<y<1$ mean, $\sigma=0<y<1$ dispersion/precision, $\nu=0$ responses, $\tau=1$ responses) were the same. The fixed effects were prescriptive grammaticality (grammatical, ungrammatical), bilingual group (late bilinguals, heritage speaker bilinguals), and the interaction of grammaticality and bilingual group. By-participant and by-item random intercepts were included as random effects\(^3\) for all parameters expect the $\sigma$ parameter due to non-convergence. A logit link function as used for the $\mu$ and $\sigma$ parameters and a log link function was used for the $\nu$ and $\tau$ parameters. The model and estimates for each parameter of the rating distribution are reported for the full

---

\(^3\) The random effects specified in all models in this dissertation only include by-participant and by-item random intercepts. No random slopes of condition by-participant or bilingual group by-item were included in any model. Barr and colleagues (2013) recommend that the maximal random effects structure that will converge be specified in a model, and Matuschek and colleagues (2017) recommend that model selection be used to determine the most maximal random effects structure supported by the data. Neither of these procedures were done as the data in reported present studies did not support inclusion of condition by-participant and/or bilingual group by-item random slopes. The items were nested within condition and participants were nested within bilingual group. Specifying random slopes in a model of this data would have controlled for the experimental effects modelled with the fixed effects. For example, we expected that the slope of condition might vary by group (greater slope, or difference between object- and subject-relatives, for late bilinguals and heritage speakers) and therefore half of the participants would have greater slope. This variation in slope is a consequent of the experimental effects and ought not to be controlled for by assuming it is inter-participant variability.
model. The estimates for the beta regression $\mu$ and $\sigma$ parameters and the estimates for the zero- and one-inflated $\nu$ and $\tau$ parameters are transformed into rating responses and the odds ratio, respectively, using the inverse of the link function.

4.3.2 Results
Mean participant-average rating is plotted in Figure 6 and in Tables Table 5-Table 9. Group differences in participant-average rating show evidence of linguistic insecurity in heritage speakers. Heritage speakers’ ratings were more towards the center than late bilinguals. In 70.00% ($n=7$) of grammatical conditions ($n=10$), heritage speakers rated grammatical sentences as less natural than late bilinguals. On grammatical items across condition (excluding TTSE items, discussed below), heritage speakers ($M=1.93$, $SD=0.76$) rated items as less natural than late bilinguals ($M=1.66$, $SD=0.65$). In 88.89% ($n=1$) of ungrammatical conditions ($n=9$), heritage speakers rated ungrammatical sentences as more natural than late bilinguals. On ungrammatical items across condition (excluding TTSE items), heritage speakers ($M=3.53$, $SD=0.79$) rated items as more natural than late bilinguals ($M=3.58$, $SD=0.87$), although this difference was not as large as it was for grammatical items. These preliminary results show evidence of a “maybe bias”. Contrary to expectation, prescriptively ungrammatical complement clause items with subject extraction (TTSE) were treated like grammatical items by both participant groups. This is likely due to influence from English and while identified as anomalous by neuroelectrical measures, the small lexical anomaly is repaired and doesn’t lead to classification as unnatural (Martohardjono et al., 2016, 2017). However, heritage speakers ($M=2.23$, $SD=0.77$) still exhibited linguistic insecurity on ungrammatical complement clause items by rating the items as less natural than late bilinguals ($M=1.95$, $SD=0.47$). Since rating data will include the main effect of grammaticality and the interaction of grammaticality with bilingual group, TTSE items are excluded from the mixed-effects model.
Figure 6: Mean participant-average naturalness ratings by condition by grammaticality by bilingual group.

Table 5: Mean participant-average naturalness rating by complex noun phrase island condition and by-and across- bilingual group. Standard deviations in parentheses.

<table>
<thead>
<tr>
<th>Island Type</th>
<th>Condition</th>
<th>Late bilinguals</th>
<th>Heritage Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grammatical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCOE</td>
<td></td>
<td>2.26 (0.49)</td>
<td>2.36 (0.73)</td>
<td>2.31 (0.63)</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>3.39 (0.65)</td>
<td>3.33 (0.65)</td>
<td>3.36 (0.65)</td>
</tr>
<tr>
<td>CCSE</td>
<td>Grammatical</td>
<td>2.01 (0.40)</td>
<td>2.06 (0.54)</td>
<td>2.04 (0.47)</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>2.96 (0.95)</td>
<td>2.83 (0.73)</td>
<td>2.89 (0.84)</td>
</tr>
</tbody>
</table>

Table 6: Mean participant-average naturalness rating by relative clause island condition and by- and across- bilingual group. Standard deviations in parentheses.

<table>
<thead>
<tr>
<th>Island Type</th>
<th>Condition</th>
<th>Late bilinguals</th>
<th>Heritage Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCSS</td>
<td>Grammatical</td>
<td>1.21 (0.25)</td>
<td>1.51 (0.56)</td>
<td>1.37 (0.47)</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>4.31 (0.44)</td>
<td>4.34 (0.34)</td>
<td>4.32 (0.39)</td>
</tr>
<tr>
<td>Island Type</td>
<td>Condition</td>
<td>Late bilinguals</td>
<td>Heritage Speakers</td>
<td>Total</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>TAOE</td>
<td>Grammatical</td>
<td>1.08</td>
<td>1.56</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.28)</td>
<td>(0.74)</td>
<td>(0.62)</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>3.75</td>
<td>3.71</td>
<td>3.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.78)</td>
<td>(0.62)</td>
<td>(0.70)</td>
</tr>
<tr>
<td>TASE</td>
<td>Grammatical</td>
<td>1.43</td>
<td>1.72</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.35)</td>
<td>(0.71)</td>
<td>(0.58)</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>3.25</td>
<td>3.53</td>
<td>3.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.82)</td>
<td>(0.81)</td>
<td>(0.82)</td>
</tr>
</tbody>
</table>

Table 7: Mean participant-average naturalness rating by temporal adverbial island condition and by- and across-bilingual group. Standard deviations in parentheses.

<table>
<thead>
<tr>
<th>Island Type</th>
<th>Condition</th>
<th>Late bilinguals</th>
<th>Heritage Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTSE</td>
<td>Grammatical</td>
<td>2.16</td>
<td>2.27</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.51)</td>
<td>(0.63)</td>
<td>(0.58)</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>1.95</td>
<td>2.23</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.47)</td>
<td>(0.77)</td>
<td>(0.66)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2.56</td>
<td>2.67</td>
<td>2.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.19)</td>
<td>(1.10)</td>
<td>(1.15)</td>
</tr>
</tbody>
</table>

Table 8: Mean participant-average naturalness rating for complement clause islands and across all condition and by- and across-bilingual group. Standard deviations in parentheses.

<table>
<thead>
<tr>
<th>Island Type</th>
<th>Condition</th>
<th>Late bilinguals</th>
<th>Heritage Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHOE</td>
<td>Grammatical</td>
<td>2.28</td>
<td>2.17</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.52)</td>
<td>(0.57)</td>
<td>(0.55)</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>3.61</td>
<td>3.52</td>
<td>3.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.61)</td>
<td>(0.51)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>WHSE</td>
<td>Grammatical</td>
<td>2.32</td>
<td>2.57</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.56)</td>
<td>(0.77)</td>
<td>(0.69)</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>2.85</td>
<td>2.94</td>
<td>2.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.87)</td>
<td>(0.77)</td>
<td>(0.81)</td>
</tr>
</tbody>
</table>

Table 9: Mean participant-average naturalness rating by wh-island condition and by- and across-bilingual group. Standard deviations in parentheses.
Table 10: Distribution of naturalness ratings by- and across- group by grammaticality.

The zero-and-one inflated beta regression estimates four parameters in Table 11 from the rating distribution in Table 10. The first two parameters of the beta regression ($\mu$ and $\sigma$) were estimated using ratings between 2-4 inclusive (41.28%, $n=2924$). For the $\mu$ (mean) parameter, the main effect of group was significant, $B_\mu=0.50$, $SE(B_\mu)=0.03$, $z=16.86$, $p<.001$. Heritage speakers rated items across condition as more unnatural and more towards the midpoint than late bilinguals. The main effect of grammaticality was also significant, $B_\mu=0.08$, $SE(B_\mu)=0.03$, $z=2.48$, $p<.05$. Ungrammatical items were rated as more unnatural than grammatical items. The interaction of group and grammaticality was not significant $B_\mu=-0.06$, $SE(B_\mu)=0.04$, $z=-1.51$, $p=.13$. Heritage speakers rated items in a similar manner as late bilinguals.

The geometric rating mean, in table 12, for heritage speakers shows that ratings for both grammatical items ($M_\mu=3.16$) and ungrammatical items ($M_\mu=3.24$) were close to the midpoint 'possible', whereas for late bilinguals, only ungrammatical ratings were close to the midpoint. For the $\sigma$ (dispersion) parameter, the main effect of group was not significant, $B_\sigma=-0.01$, $SE(B_\sigma)=0.05$, $z=-0.11$, $p=.91$. Only the main effect
of grammaticality was significant, \( B_\sigma = 0.22, SE(B_\sigma) = 0.05, z = 4.33, p < .001 \). The interaction of group and grammaticality was also not significant, \( B_\sigma = 0.07, SE(B_\sigma) = 0.07, z = 0.99, p = .32 \). There was significantly more dispersion for ungrammatical ratings than for grammatical ratings across group. The geometric rating mean, in Table 12, for ungrammatical items (\( M_\sigma = 2.18 \)) is larger than for grammatical items (\( M_\sigma = 2.00 \)). This indicates more uncertainty in ungrammatical ratings in a similar way for each of the bilingual groups.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fixed Effect</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>(Intercept)</td>
<td>-0.34</td>
<td>0.02</td>
<td>-14.75</td>
<td>&lt; .001 ***</td>
</tr>
<tr>
<td></td>
<td>( \text{link=logit} )</td>
<td>Group(HS)</td>
<td>0.50</td>
<td>0.03</td>
<td>16.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cond(Ungram)</td>
<td>0.08</td>
<td>0.03</td>
<td>2.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group:Cond</td>
<td>-0.06</td>
<td>0.04</td>
<td>-1.51</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>(Intercept)</td>
<td>-1.09</td>
<td>0.04</td>
<td>-28.56</td>
<td>&lt; .001 ***</td>
</tr>
<tr>
<td></td>
<td>( \text{link=logit} )</td>
<td>Group(HS)</td>
<td>-0.01</td>
<td>0.05</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cond(Ungram)</td>
<td>0.22</td>
<td>0.05</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group:Cond</td>
<td>0.07</td>
<td>0.07</td>
<td>0.99</td>
</tr>
<tr>
<td>( \nu )</td>
<td>(Intercept)</td>
<td>1.12</td>
<td>0.07</td>
<td>15.92</td>
<td>&lt; .001 ***</td>
</tr>
<tr>
<td></td>
<td>( \text{link=log} )</td>
<td>Group(HS)</td>
<td>-3.29</td>
<td>0.12</td>
<td>-27.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cond(Ungram)</td>
<td>-0.58</td>
<td>0.10</td>
<td>-6.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group:Cond</td>
<td>0.22</td>
<td>0.17</td>
<td>1.33</td>
</tr>
<tr>
<td>( \tau )</td>
<td>(Intercept)</td>
<td>-3.38</td>
<td>0.16</td>
<td>-20.84</td>
<td>&lt; .001 ***</td>
</tr>
<tr>
<td></td>
<td>( \text{link=log} )</td>
<td>Group(HS)</td>
<td>3.28</td>
<td>0.18</td>
<td>18.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cond(Ungram)</td>
<td>0.73</td>
<td>0.20</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group:Cond</td>
<td>-1.09</td>
<td>0.22</td>
<td>-4.97</td>
</tr>
</tbody>
</table>

Table 11: Maximal zero-and-one inflated beta regression mixed-effects model coefficients by rating distribution parameters.

The \( \nu \) parameter was estimated using rating of 1 ‘es natural’ (37.78%, \( n = 2676 \)). For the \( \nu \) parameter (ratings of natural), the main effect of group was significant, \( B_\nu = -3.29, SE(B_\nu) = 0.12, z = -27.34, p < .001 \). Heritage speakers were 0.04 times as likely as late bilinguals to rate items as fully natural. The main effect of grammaticality was also significant, \( B_\nu = -0.58, SE(B_\nu) = 0.10, z = -6.11, p < .001 \). Participants were more than half as likely (\( OR = 0.56 \)) to rate ungrammatical items as fully natural. The interaction of group and grammaticality was not significant (\( B_\nu = 0.22, SE(B_\nu) = 0.17, z = 1.33, p = .19 \)) indicating that heritage speakers patterned similarly in the way they rated grammatical versus ungrammatical items as fully
natural. The odds ratios, in Table 13, show that late bilinguals are more likely to use the ‘es natural’ endpoint for grammatical ($M_{\nu}=3.07$) and ungrammatical items ($M_{\nu}=1.72$), although less than grammatical items, than heritage speakers are (grammatical: $M_{\nu}=0.11$; ungrammatical $M_{\nu}=0.06$).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Group</th>
<th>Condition</th>
<th>Estimate</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>Late bilinguals</td>
<td>Grammatic</td>
<td>-0.34</td>
<td>2.66</td>
</tr>
<tr>
<td></td>
<td>link=logit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heritage Speakers</td>
<td>Grammatic</td>
<td>0.16</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td>link=logit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Across</td>
<td>Grammatic</td>
<td>-1.09</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>link=logit</td>
<td></td>
<td></td>
<td>2.18</td>
</tr>
</tbody>
</table>

Table 12: Beta regression estimates and geometric rating means by group and by condition.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Group</th>
<th>Condition</th>
<th>Estimate</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu$</td>
<td>Late bilinguals</td>
<td>Grammatic</td>
<td>1.12</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>link=log</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heritage Speakers</td>
<td>Grammatic</td>
<td>-2.17</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>link=log</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>Late bilinguals</td>
<td>Grammatic</td>
<td>-3.38</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>link=log</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heritage Speakers</td>
<td>Grammatic</td>
<td>-0.09</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Table 13: Beta regression estimates and odds ratio by group and by condition.

The $\tau$ parameter was estimated using rating of 5 'no es natural' (20.95%, $n=1484$). For the $\tau$ parameter (ratings of unnatural), the main effect of group was significant, $B_{\tau}=3.28$, $SE(B_{\tau})=0.18$, $z=18.74$, $p<.001$. Heritage speakers were 26.58 times as likely as late bilinguals to rate items as fully unnatural. The main effect of grammaticality was significant, $B_{\tau}=0.73$, $SE(B_{\tau})=0.20$, $z=3.68$, $p<.001$. Late bilinguals were 2.08 times as likely to rate ungrammatical items as fully unnatural. The interaction of group and grammaticality was significant, ($B_{\tau}=-1.09$, $SE(B_{\tau})=0.22$, $z=-4.97$, $p<.001$), indicating that heritage speakers patterned different from late bilinguals with respect to the difference between grammatical and ungrammatical items. The odds ratios, in Table 13, show that late bilinguals are much less likely to use the ‘no es natural’ endpoint for ungrammatical ($M_{\nu}=0.07$) and grammatical items ($M_{\nu}=0.03$), although less so for grammatical items, than heritage speakers are (grammatical: $M_{\nu}=0.91$; ungrammatical $M_{\nu}=0.64$). This is
counterintuitive and possibly represents more tolerance of ungrammaticality in terms of how it impacts naturalness for late bilinguals.

4.3.3 Discussion

The analysis of naturalness ratings of syntactically complex Spanish sentences showed that heritage speakers do indeed rate items differently than late bilinguals, which can be argued to reflect linguistic insecurity when completing a task requiring metalinguistic judgments. Across nine different conditions, heritage speakers were less likely to find a sentence fully natural and used the mid-point more than late bilinguals. In the present study, the midpoint of the Likert scale was labelled es posible ‘maybe’ consequently, the linguistic insecurity of heritage speakers manifests as failure to fully accept or fully reject most items, and a bias towards rejecting items (including ungrammatical items). It is unclear if this “maybe-bias” is a function of heritage speakers’ linguistic insecurity or common edge-avoidance or “fence-sitting”. It has been argued that insecurity might lead heritage speakers to over-accept items as a “yes-bias”, however, the findings suggest that linguistic insecurity leads heritage speakers to not feel confident in deciding something is ‘good’ but can still identify ‘bad’ items, contra the predictions of a heritage speaker “yes-bias”.

Late bilinguals, however, did not respond as one might expect given the prescriptive grammaticality of the sentences. Overall, late bilinguals tended to be more accepting of items than heritage speakers. Late bilinguals were more likely to rate an item as fully natural, and less likely to rate an item as fully unnatural. The pattern of late bilinguals being more accepting of ungrammatical items is similar to results from teachers’ ratings of errors in L2 student writing. Native English-speaking teachers were more lenient of student errors than non-native English speaking teachers (e.g., Majer 1983; Hyland & Anan, 2006). Likewise, the pattern of heritage speakers being more likely to fully reject items is consistent with findings that individuals with high linguistic insecurity are more sensitive to norms (Labov, 1990). The findings of the present study confirm the claims that metalinguistic tasks are inappropriate for heritage speakers, but also suggest that these tasks are inappropriate for late bilinguals as well. An additional interesting finding is that there was greater consensus in rating grammatical items than ungrammatical items. This suggests that while notions of naturalness or grammaticality converge, the
relationship between naturalness and ungrammaticality is less clear. This could result from the fact that ungrammatical sentences have no underlying cognitive representation (Sprouse, 2007b) and therefore, naturalness ratings of ungrammatical items have no comparison point.

5. Heritage speaker summary
Heritage speakers are grossly understudied, yet prevalent, members of the linguistic community in the Global North and more research is needed. Compared to traditional models of a speaker/listener in generative linguistics, they are very different from monolinguals and late adult L2 learners, the typical foci of multilingualism research. Heritage speakers are raised, educated, and work in an ML dominant society that results in them using the HL and ML frequently in different settings, but having similar language abilities to their parents, late bilinguals, in the HL. Previous research has focused on deficit models of heritage speaker language ability and linguistic knowledge by inappropriately comparing them to monolinguals educated in the standard, prestige varieties of the HL using inappropriate measures and statistical analyses.

6. Heritage speaker processing
The present study looks not at (metalinguistic) grammatical knowledge or production, but rather at language processing. The difference is that the underlying representation in the mind of a speaker cannot be accessed by asking people what they think about grammar (metalinguistic grammaticality judgment tasks) or through studies of language production (see below for discussion). The subject of research in generative linguistics is the speaker/listener’s linguistic competence, not their performance, which is subject to numerous confounds. It is unlikely that one can ever measure the psychological construct of linguistic competence absent performance. Even metalinguistic grammaticality tasks fundamentally tap performance since the speaker/listener must reflect on language. The best option is to measure the processing of language since it removes conscious reflection of the speaker/listener and avoids the obvious confounds of production albeit with unavoidable confounds such as audition.

This section briefly outlines problems with production tasks, grammaticality judgment tasks, and deficit models of heritage speaker grammar. As mentioned above, English is the default language of
learning and teaching from pre-school through college (Beaudrie & Fairclough, 2012; Valdes, Fishman, Chavez & Perez, 2006), so heritage speakers of Spanish in the anglophone US are generally not given the opportunity to acquire literacy skills or overt metalinguistic instruction in Spanish. Martohardjono, Valian, & Klein (2018) argue that the linguistic competence of a language user cannot be explained by errors in production. Oral production errors are not a valid measure of syntactic representation because they cannot be separated from effects of performance, including discourse and message planning, lexical look-up, phonetic/phonological output mechanisms, and the integration of those systems. Production studies cannot distinguish between the results of syntactic processes and performance, or post-syntactic processes, and therefore, they cannot be representative of the mental grammar. To avoid the confound caused by metalinguistic tasks that are biased toward populations that are educated in the language under investigation, implicit measures of processing are preferable (Kotz, 2009; Moreno, Rodriguez-Fornells & Laine, 2008).

The majority of literature on heritage speakers focuses not on language processing, but attempts to make claims about heritage speaker linguistic competence. The focus of the research has been on identifying points where the HL and ML differ to probe heritage speakers' grammar, not on processing patterns in their dominant and non-dominant language. The research conducted under the consensus view of heritage speakers as having attrited, incomplete, or simpler grammars of the HL as the cause for the differences between monolingual and heritage speaker grammars (See Benmamoun, Montrul, & Polinsky (2010)) has not provided any empirical evidence that these differences are not a result of processing differences in heritage speakers. Indeed, differences between the psychological construct of grammar may not be provable in a heritage speaker population.

In the next section, theories of bilingual processing are reviewed and discussed in the context of heritage speakers. Heritage speaker dominance vs. first learnt language must be considered, as the heritage speaker situation is different from late bilinguals, which most of the research is conducted on. Unfortunately, a coherent theory of language processing that serves heritage speakers is not available, and it is outside the scope of this dissertation to create one. Because there is so little research available on heritage speaker processing, the goal for this investigation is to provide empirical evidence that will contribute to the base upon which a theory of bilingual processing will draw.
6.1 Major models of bilingual language processing

The focus of language processing research in multilinguals has been conducted on late L2 adult learner populations. Proposed theories accounting for multilingual language processing therefore are argued based on evidence principally from a population fundamentally different from heritage speaker bilinguals and additionally from a bilingual deficit framework. Two major theories of bilingual language processing are presented and discussed below. Their applicability to heritage speaker language processing is discussed.

6.1.1 Shallow processing

The Shallow Structure Hypothesis is a model showing the representation of the grammar based on the Continuity of Parsing Hypothesis (Clahsen & Felser, 2006). The Continuity of Parsing Hypothesis claims that L1 processing mechanisms for morphology and syntax are the same for adults and children (there is empirical evidence for this from children; though all children in such studies are over six years old.) Therefore, under this model, L1 processing mechanisms do not develop or change over time, and any difference in child processing as compared to adult processing is due to cognitive and working memory differences.

The Shallow Structure Hypothesis applies this finding to late L2 learners with the claim that L2 speakers do not use syntactic information to process complex L2 grammatical structures (Keating, 2009; Marinis et al., 2005; Papadopoulou & Clahsen, 2003; Sato & Felser, 2006). According to the Shallow Structure Hypothesis, native speakers have two ways in which they are able to parse sentences. One is based on deeper level syntactic processing, and the other is based on surface level use of semantic and pragmatic context. However, this model claims that L2 learners have only the lexical, semantic, and pragmatic levels available to them, and are unable to access L2 grammatical representations. This model is based on the assumption that L2 learners cannot make use of the mental representation to process utterances, because their grammatical representation is constrained by critical period effects which blocks the development of a full grammatical representation.
Zhang (2017) evaluates the Shallow Structure Hypothesis and the Continuity of Parsing Hypothesis in the context of heritage language speakers. A confound occurs because studies showing the Continuity of Parsing Hypothesis are based on children over six, and heritage speakers may experience a decrease in input from the home language before the age of six. It is therefore possible that a child will experience a change of language input during early HL acquisition. The effect of this change on the language processing mechanisms is unknown, and at the present time, a control group of children under the age of six who have not experienced a change in the language they receive input in is not available. If heritage speakers and monolinguals have similar processing patterns, and if the pattern observed for children over age six holds true for children under the age of six, then it is possible there is indeed no change to the processing mechanisms. However, according to Zhang, if changes in processing patterns for heritage speakers are observed, than the Continuity of Parsing Hypothesis would be falsified.

6.1.2 Declarative-procedural

The Declarative-Procedural Model is a model of processing mechanisms (Ullman, 2004). This model is not concerned with grammatical representation, but rather mechanisms of access to knowledge. The model assumes a declarative memory system, which is activated in the acquisition and processing of facts and events; and a procedural memory system, which is activated in the acquisition and use of unconscious knowledge and skills, like cognition and seemingly automatic processes. The Declarative-Procedural Model is applied to language by separating language components into the mental lexicon and the mental grammar. The mental lexicon is part of the declarative memory system, and is responsible for lexical storage, including lexical form and meaning, irregular morphological forms, and other memorized forms. The mental grammar is responsible for hierarchical linguistic structures, including syntax, regular morphological forms, and other implicit language knowledge.

This model states that L2 learners are able to build and access the declarative memory system, which includes the mental lexicon, and accounts for the observation that L2 learners begin by memorizing words and templatic structures rather than building sentences structurally. The Declarative-Procedural Model also states that late L2 learners are not able to access procedural memory as they begin to learn the L2. However, this model accounts for the change that comes with increased frequency of use of a L2
and increasing language proficiency by proposing that increased frequency of use can also increase the ability to access procedural memory. If an L2 learner becomes highly proficient or even dominant in their L2, the ability of an L2 learner can go so far as to converge with the nativelike way the brain processes L1, automatizing grammatical processes in procedural memory. This theory is attractive because it is representative of the different range of skills and abilities of bilinguals. However, there are some limitations (see Brill-Schuetz & Morgan-Short, 2014; Morgan-Short et al., 2014). The model does not provide a mechanism by which knowledge can transfer from declarative to procedural memory and does not account for why some learners are able to access procedural memory faster or easier than others.

Zhang (2017) refutes the use of the Declarative-Procedural Model to explain heritage speaker language use, citing Bolger and Zapata’s (2011) discussion of how the Declarative-Procedural Model would apply to heritage speakers. Zhang argues that their model cannot account for why or how heritage speakers process some grammatical structures differently from monolinguals.

6.2 Heritage speakers and models of processing

Both the Shallow Structure Hypothesis and Declarative-Procedural Model were developed for late L2 adult bilingual learners, and are framed in terms of a bilingual deficit. As discussed earlier in the section, extensions of the models are typically based on the inappropriate comparison of heritage language speakers to monolingual speakers of a prestige variety. Proving or disproving a theory of bilingual processing is beyond the scope of this dissertation, as is proposing a new model of heritage speaker processing. An appropriate model of heritage language processing would have predicted differences between heritage speaker processing, late bilingual processing, and monolingual processing. Furthermore, in order for the model to be sufficient, factors that explain the differences between heritage language processing, late bilingual processing, and monolingual processing, must be explored, as well as the bidirectional influence of all of a bilingual’s languages. See Zhang (2017) for further discussion of models of processing for heritage speakers.
7. Summary

In this chapter, the concept of heritage speakers has been discussed and confounds unique to their research has been reviewed and motivated. Results from a principal component analysis provide evidence, in line with previous research, that heritage speakers utilize the ML in public settings and are fluent and frequent users of both the HL and ML, similar to the appropriate comparison group: late adult L2 learners in the same community in an L2 dominant environment. Previous research has nearly exclusively discussed heritage speakers as being deficient and incomplete speakers of their HL, however, these claims are called into question by the confounds of linguistic insecurity and the inappropriateness of the measures used to establish the deficit framework. It is peculiar that underperformance on a task can be both diagnostic of incomplete acquisition or lack of control of a grammatical construct, and also evidence of extra-linguistic insecurity. Results from a novel application of a relatively new statistical technique more appropriate to acceptability rating task data suggests that heritage speakers are more critical in their HL than late bilinguals. Lastly, the focus on language processing to the exception of grammatical knowledge was motivated due to a myriad of unavoidable confounds. Models of bilingual language processing were considered and found to be inappropriate for heritage speaker language processing.
CHAPTER III:
THE PRESENT STUDY

The present study examines heritage speaker sentence processing to address the issues presented in chapter two using appropriate research methodologies and an appropriate comparison group. Heritage speakers will not be compared to the usual inappropriate monolingual baseline, but rather other bilingual speakers of the HL living in a ML dominant environment. Late bilinguals are the comparison group for this study. They are Spanish speakers raised in Spanish-dominant societies (i.e., Spanish-speaking North America, Central America, South America, the Caribbean, Puerto Rico) who emigrated to the US after age 18. Late bilinguals serve as the time-apparent parents of heritage speakers and similarly work and socialize in an ML dominant society. The appropriate methodologies used are implicit, do not require metalinguistic reflection and decisions, and use stimuli presented in a manner appropriate for bilinguals not formally educated in the HL. Aurally presented stimuli spoken by a Spanish L1 speaker avoid the confound of literacy common in other studies.

The data come from the ongoing multi-year multi-methodological Second Generation Bilinguals Project conducted jointly by the Second Language Acquisition Lab directed by Gita Martohardjono, the Child Language Lab directed by Richard G. Schwartz, and the Research Institute for the Study of Language in Urban Society co-directed by Gita Martohardjono and Ricardo Otheguy at the Graduate School and University Center of the City University of New York. The Second Generation Bilinguals Project utilizes eye tracking, pupillometry, event-related potentials, acceptability judgements, picture-pointing comprehension, and survey tasks in the study of Spanish heritage speakers and late bilinguals in the greater New York City area. The project focuses on intergenerational language change in the bilingual Latinx community in New York City and syntactic processing of wh-extraction from islands. Data analyzed and presented in this dissertation have not been previously utilized by the project in its conference presentations and publications.

In this chapter, the present study will be motived and outlined. Section one will introduce the psychophysiological research protocols used in the present study. Section two will introduce and motivate the subject-object relative clause processing asymmetry. The robustness of this phenomenon makes it an
ideal candidate for investigation. Section three will present previous heritage speaker studies of the asymmetry and present the results of a new comprehension study of heritage speaker accuracy on relative clauses. Lastly, section four will introduce the main research questions and hypotheses of the dissertation and section five will conclude.

1. Methodology rationale

As demonstrated and discussed in the previous chapter, grammaticality judgement tasks and production tasks are not suitable for investigating heritage speaker language processing. An appropriate measure: 1) does not require metalinguistic judgments, 2) must measure automatic preconscious reflexes to language processing, and 3) excludes written presentation of stimuli. Given these constraints, the ideal measures are psychophysiological measures, that is, methodologies that directly measure the body's automatic physiological responses to psychological phenomena, such as language. In language processing, and with any cognitive activity, increased activation of cortical neural cells is observed. When activated, neural cells receive, process, and transmit information through a combination of chemical and electrical signals. This electrical cortical activity can be indirectly measured by recording changes in the electrical or magnetic field around the brain. The chemical cortical activity can be indirectly measured by recording physiological changes concomitant with the increase in neurotransmitters. The present study will utilize two indirect measures of the brain's activation in language processing: event-related potentials and pupillometry.

1.1 Event-related potentials

The most frequently employed psychophysiological measure in linguistics research utilizes event-related potentials (ERPs). ERPs are neuroelectrical recordings, electroencephalograms, taken at the scalp of small voltage changes that result from the presentation of some event, such as processing an anomalous or ungrammatical sentence. By recording this neuroelectrical activity in response to an event, we can capture signatures of online language processing in a temporally fine-grained way. ERPs have been used to study language processing in monolinguals (e.g., Kutas & Hillyard, 1980; Osterhout, 1994; Friederici, 2002), bilinguals (e.g., Moreno & Kutas, 2005; Tokowicz & MacWhinney, 2005; Sabourin et al., 2006;
Moreno, Rodríguez-Fornells, & Laine, 2008; Kotz, 2009; Dowens et al., 2009), and trilinguals (e.g., Aparicio et al., 2012; Grey, Sanz, Morgan-Short, & Ullman, 2017) and provide a well-documented measure of processing difficulty (Osterhout, 1994). Changes in neuroelectrical signatures as a result of the acquisition and learning of L2 features have been described in longitudinal (Osterhout et al., 2006; Osterhout et al., 2008; McLaughlin et al., 2010; White, Genesse, & Steinhauer, 2012) and cross-sectional studies of language learners (Gabriele, Fiorentino, & Alemán Bañón, 2013; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013), as well as after instruction (Friederici, Steinhauer, & Pfeifer, 2002; McLaughlin, Osterhout, & Kim, 2004; Mueller et al., 2005; Davidson & Indefrey, 2008; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010).

ERPs as a measure of language processing has not been shown to be sensitive to feelings of linguistic insecurity or indeterminacy, which are common in heritage speakers (Benmamoun et al., 2010; Martinez & Petrucci, 2009) and are exacerbated by notions of correctness that drive metalinguistic judgments. The research protocol is an ideal measure to use for bilinguals as it does not require action by the participant and the neurological responses recorded are automatic. Additionally, the stimuli in this ERP experiment are presented as natural, running speech and are particularly well suited for studying processing of the first-learnt language of heritage speakers as it avoids the confounds of literacy, education, and metalinguistic reflection.

1.2 Pupillometry

It is well known that the pupil dilates in response to changes in light. Over 250 years ago, pupillary responses to fear or excitement were observed in animals (Fontana, 1765, as cited in Loewenfeld, 1958). Research in the late 1900s extended the observation of non-light related pupillary reflexes to include pupillary dilations in response to multiplication tasks (Heinrich, 1986, as cited in Bourisly, 2015). Since the 1960s (e.g., Kahneman & Beatty, 1996) there has been an exponential increase in studies exploring pupillary responses to a variety of cognitive tasks (see Beatty, 1982; Beatty and Lucero-Wagoner 2000 for thorough reviews of research). Therefore, the second measure utilized in the present study is pupillometry.
Pupillometry is an under-utilized implicit measure of cortical activity. The increase of the neurotransmitter norepinephrine (or noradrenaline) by activation of the locus coeruleus area of the brain facilitates cognitive processing. Noradrenaline is a neurotransmitter that has a role in attention, sensory processing, memory formation and retrieval, decision making and performance facilitation, among others (Sara, 2009). Activity in the locus coeruleus has been linked to the P3 ERP components seen in a number of language studies (Nieuwenhuis, Aston-Jones, & Cohen, 2005). Increases in noradrenaline result in activation of the sympathetic division of the autonomic nervous system. The stimulation of the sympathetic nervous system results in pupil dilation in humans (McDougal & Gamlin, 2015) as well as monkeys (Joshi et al., 2016), stimulated sweat production increasing skin conductivity (Kintsch, 1965), vasoconstriction (Elias & White, 1969), and increased heart rate. Autonomic responses such as noradrenaline and locus coeruleus activity is likely involved in cognitive load and language processing as they facilitate neuronal synchronization (Demberg & Sayeed, 2016). These physiological changes have been utilized in linguistic research using galvanic skin response or skin conductance/impedance tasks and crucially in pupil diameter change tasks. Increased cortical activity facilitated by norepinephrine can then be measured as changes in pupil dilation. Increases in pupil dilation have been varyingly argued to index processing load (e.g., Ahern & Beatty, 1979; Beatty, 1982; Granholm et al., 1996), memory load (e.g., Beatty, 1982; van Rijn et al., 2012), cognitive load (e.g., Hess & Polt, 1964; Goldwater, 1972; Klinger, 2010; Bourisly, 2015), and capacity utilization (Just, Carpenter, & Miyake, 2003).

Measurement of the pupil diameter, referred to as pupillometry, is useful as a non-invasive technique that require no conscious action on the part of the participant. Pupillometry records near real time responses to stimuli, referred to as task-evoked pupillary responses (TEPR), are similar to ERP albeit slower and have fewer restrictions in terms of stimuli design. Similar to electroencephalographic recordings, pupil diameter measures allow for auditory presentation and measure the automatic cortical activity associated with language processing. In this study, the implicit nature of the measure is well suited to study the processing of the first-learned language of heritage speakers as it avoids the confounds of literacy, education, and metalinguistic reflection, similar to ERP.
2. Relative clause processing asymmetry

To study the language processing of heritage speakers appropriately, an appropriate linguistic phenomenon is needed. The parameters of such a language phenomenon are: 1) not susceptible to cross-linguistic influence or transfer effects, 2) not susceptible to attrition or simplification, and 3) a robust, well-established phenomenon. Given these constraints, an ideal choice is the subject-object relative clause processing asymmetry. The subject-object relative clause processing asymmetry in Spanish-English bilinguals is ideal as the phenomenon is not susceptible to transfer from the dominant language, English, in heritage speakers. Researchers have attributed differences in heritage speaker’s grammars to transfer effects (Montrul, 2010; Scontras, Fuchs, & Polinsky, 2015) and may be a confound in investigations of heritage speaker language processing. The syntax of relative clauses is similar in Spanish (the HL) and English (the ML) obviating this confound. The subject-object relative clause processing asymmetry, is likely not susceptible to attrition since relative clauses are early acquired and argued to be cross-linguistically universal. This is similar to the syntactic island constraint phenomenon which has been argued to be cross-linguistically invariable and “largely immune to environmental influences and stem from deeper properties of the processor and/or grammar” based on evidence from Korean heritage speakers (Kim & Goodall, 2016, p. 3). The subject-object relative clause processing asymmetry is also a robust, phenomenon that is well-established across a wide variety of languages and in a large number of studies employing a myriad of different methodologies.

In the following sections, relative clauses in Spanish and English will be introduced. Evidence of the subject-object relative clause processing asymmetry is reviewed in acquisition studies and as revealed by various psycholinguistic methods. A review of the cross-linguistic evidence for the asymmetry is also presented. Sources of the subject-object relative clause processing asymmetry will also be reviewed, although it is outside the scope of the present study to evaluate these sources.

2.1 Subject-gap relative clauses vs object-gap relative clauses

A relative clause is a subordinate clause embedded within a nominal phrase, which modifies a head noun. It is an example of a long-distance dependency consisting of a nominal head (the modified element) and a relativized complementizer phrase (the dependent element). In Spanish, as well as in
English, the nominal head is associated with an empty element, or gap position, within the subordinate clause. The relative clause in Spanish is marked with an obligatory relativizer (i.e., que or quien when the nominal head is human). Spanish nominally has subject-verb-object (SVO) word order but does allow for post-verbal subjects. In subordinate clauses, the typical, but not obligatory, word order is verb-subject (Gutiérrez-Bravo, 2003). Relative clauses vary in whether the nominal head is associated with a gap in the subject position or object position of the subordinate clause. First and second acquisition literature emphasizes the difference between subject-gap relative clauses and object-gap relative clauses. A subject-gap relative clause such as example (11) below, has a noun phrase Bertha which is the subject of both clauses: the matrix clause Bertha viajó por europa ‘Bertha traveled to Europe’ and the subordinate clause Bertha ama a Little ‘Bertha loves Little’. An object-gap relative clause such as example (12) below, has a noun phrase Bertha which is the subject of the main clause Bertha viajó por europa ‘Bertha traveled to Europe’, but the object of the subordinate clause Little ama a Bertha ‘Little loves Bertha. Note that in object-gap relative clauses with a human head noun, the relativizer quien is marked with a, a direct object marker. In both the subject-gap (11) and object-gap (12) relative clause, the relativized noun phrase is not expressed in the subordinate clause and indicated with a gap __.

Subject-gap relative

(11) Bertha [quien __ ama a Little] viajó por europa.
‘Bertha who loves Little traveled to Europe’

Object-gap relative

(12) Bertha [a quien Little ama __] viajó por europa.
‘Bertha who Little loves traveled to Europe’

Embeddedness is another factor in relative clauses. Embeddedness is the position of the relative clause as the subject or the object of a sentence. In a subject-embedded relative clause such as example (13) below, the subordinate clause Silvia besó a Chris ‘Silvia kissed Chris’ is the subject of the matrix clause Silvia chismeó con Sandra ‘Silvia gossiped with Sandra’. In an object-embedded relative clause such as example (14) below, the subordinate clause Chris chismeó con Sandra ‘Chris gossiped with Sandra’ is the object of the matrix clause Silvia besó a Chris ‘Silvia kissed Chris’.
Subject-embedded relative
(13) Silvia [quien __ besó a Chris] chismeó con Sandra.
‘Silvia who kissed Chris gossiped with Sandra’

Object-embedded relative
(14) Silvia besó a Chris [quien __ chismeó con Sandra].
‘Silvia kissed Chris who gossiped with Sandra’

These two axes create a four-way distinction of relative clauses. The four types of relative clauses are as follows: subject-subject relative clauses are subject-embedded with subject gaps, as in (15); subject-object relative clauses are subject-embedded with object gaps, as in (16); object-subject relative clauses are object-embedded with subject gaps, as in (17); and object-object relative clauses are object-embedded with objects gaps, as in (18) (de Villiers, Flusberg, Hakuta & Cohen, 1979). In the subject-subject relative clause in (15), Jorge is both the subject of the matrix clause Jorge limpia la casa ‘Jorge cleans the house’ and the subordinate clause Jorge adora a Carlos ‘Jorge loves Carlos’. In the subject-object relative clause in (16), Jorge is the subject of the matrix clause Jorge limpia la casa ‘Jorge cleans the house’ and object of the subordinate clause Carlos adora a Jorge ‘Carlos loves Jorge. In the object-subject relative clause in (17), Carlos is the object of the matrix clause Jorge adora a Carlos ‘Jorge loves Carlos’ and subject of the subordinate clause Carlos limpia la casa ‘Carlos cleans the house’. In the object-object relative clause in (18), Carlos is the object of the matrix clause Jorge adora a Carlos ‘Jorge loves Carlos’ and object of the subordinate clause Jesús llamó a Carlos ‘Jesús called Carlos’.

Subject-subject relative
(15) Jorge [quien __ adora a Carlos] limpia la casa.
‘Jorge who loved Carlos cleans the house’

Subject-object relative
(16) Jorge [a quien Carlos adora __] limpia la casa.
‘Jorge who Carlos loves cleans the house’

Object-subject relative
(17) Jorge adora a Carlos [quien __ limpia la casa].
‘Jorge loves Carlos who cleans the house’

Object-subject relative
(18) Jorge adora a Carlos [a quien Jesús llamó __].
‘Jorge loves Carlos who Jesús called’
English relative clauses share a similar structure with Spanish. The two biggest differences between relative clauses in English and Spanish is word order and complementizer use. Word order in English is fixed as subject-verb-object under most circumstances in both the matrix and subordinate clauses. While the Spanish complementizer que or quien is obligatory in all four types of relative clauses above, the English complementizer that or who is optional in object-gap (subject-object and object-object) relative clauses (Sánchez-Walker & Montrul, 2016). The English examples in (19)-(22) show a very similar structure to the Spanish examples in (15)-(18).

Subject-subject relative

Subject-object relative

Object-subject relative
(21) Ariana texted Briana [who __ visited Jay].

Object-subject relative
(22) Ariana texted Briana [who Jay visited __].

2.2 Subject-Object relative clause asymmetry

There is an asymmetry in the processing of subject-gap and object-gap relative clauses. Depending on the language, one of the relative clause gap-types incurs more processing costs and the other is easier, referred to as a preference. The majority of languages examined have provided evidence for a subject-gap preference, meaning that object-gap relatives are harder to process. Evidence for this comes from acquisition literature (subject-gap relative clauses are acquired earlier in the L1 and L2). Additional evidence also comes from various methodologies commonly employed in psycholinguistic research (subject-gap relative clauses are faster to process or processed with less cortical activity) demonstrating that the phenomenon is not an artifact of one research methodology. Evidence for the asymmetry comes from a variety of languages as either a subject-gap or object-gap preference, but a preference or asymmetry nonetheless. Sources of the asymmetry have been ascribed to additional burdens placed on working memory, reanalysis by the parser, or surprisal that result in processing the less-preferred relative clause gap-type. The evidence for and potential sources of the asymmetry are presented and discussed more in the following sections.
2.2.1 Acquisition evidence of asymmetry

Relative clauses are acquired relatively early, beginning around ages two to three (McKee, McDaniel, & Snedeker, 1998), but are not fully understood by children even at four or five years of age (de Villiers & de Villiers, 1985). A consistent pattern in both L1 and L2 acquisition of English is that subject-gap relative clauses are produced and understood to a greater degree than object-gap relative clauses (Brown, 1971; Smith, 1974; de Villiers, Flusberg, Hakuta, & Cohen, 1979). There is theoretical support for the following acquisition order (with earlier acquired structures followed by later acquired structures) object-subject > subject-subject > subject-object > object-object. Subject gaps are much easier for children to understand than object gaps, and that given a particular subject gap or object gap, object-embedded sentences are acquired earlier than subject-embedded sentences (DeVilliers, Flusberg, Hakuta, & Cohen, 1979).

DeVilliers and colleagues (1979) found experimental support in an act out comprehension task administered to children for better performance on object-subject and subject-subject relative clauses, followed by object-object relative clauses and then subject-object relative clauses. This hierarchy also extends from children to L2 learners, and is a widely documented phenomenon (Gibson, 1998; Reali & Christiansen, 2007).

2.2.2 Experimental evidence of asymmetry

The preference of subject-gap relative clauses over object-gap relative clauses is evident across various processing methodologies, listed in Table 14, including behavioral, ocular, and neurological measures. Behavioral studies, where increased processing difficulty with object-gap relatives is indexed by increased reading time (King & Just, 1991; Gibson, Desmet, Grodner, Watson, & Ko, 2005), increased lexical decision time (Ford, 1983), and lower accuracy on comprehension measures (e.g., Polinsky, 2011). Ocular studies found increased reading time (Traxler, Morris, & Seely, 2002) and increased pupil dilation (Just & Carpenter, 1993; Piquado, Isaacowitz, & Wingfield, 2010) by tracking eye-gaze and pupil diameter when reading or listening to object-relatives. Neurological measures have found increased electrical amplitude in ERP studies (King & Kutas, 1995), increased blood oxygenation in functional magnetic resonance imaging (fMRI) studies (Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Caplan et al., 1999, 2000, 2001, 2002; Cooke et al., 2001, 2002) and increased glucose metabolism in positron
emission tomography (PET) studies (Stromswold, Caplan, Alpert, & Rauch, 1996; Caplan et al., 1998, 1999, 2000) indexing increased cognitive-linguistic processing associated with object-gap relative clauses in comparison to subject-gap relative clauses.

<table>
<thead>
<tr>
<th>Study type</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Behavioral</strong></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>Polinsky, 2011</td>
</tr>
<tr>
<td>Reading Time</td>
<td>King &amp; Just, 1991; Gibson, Desmet, Grodner, Watson, &amp; Ko, 2005</td>
</tr>
<tr>
<td>Lexical Decision</td>
<td>Ford, 1983</td>
</tr>
<tr>
<td><strong>Ocular</strong></td>
<td></td>
</tr>
<tr>
<td>Eye Tracking</td>
<td>Traxler, Morris, &amp; Seely, 2002</td>
</tr>
<tr>
<td>Pupillometry</td>
<td>Just &amp; Carpenter, 1993; Piquado, Isaacowitz, &amp; Wingfield, 2010</td>
</tr>
<tr>
<td><strong>Neurological</strong></td>
<td></td>
</tr>
<tr>
<td>ERP</td>
<td>King &amp; Kutas, 1995</td>
</tr>
<tr>
<td>PET</td>
<td>Stromswold, Caplan, Alpert, &amp; Rauch, 1996; Caplan et al., 1998, 1999, 2000</td>
</tr>
</tbody>
</table>

Table 14: Experimental evidence of subject-object relative clause processing asymmetry.

Figure 7: Languages showing evidence of subject-object processing asymmetry
2.2.3 Crosslinguistic evidence of asymmetry

The subject-object relative clause processing asymmetry has been shown to be a robust phenomenon cross-linguistically. It has been supported in at least twenty-seven over- and under-studied languages from twelve typologically diverse language families and isolates across all populated continents except for Africa listed in Tables Table 15 and Table 16 and illustrated in Figure 7. Map was made using the rworldmap package in R (South, 2011). The majority of studies have found evidence of a subject-gap relative clause preference, although a handful have also found evidence for an object-gap, however, in either case, a processing asymmetry has been observed.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Language</th>
<th>References</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basque</td>
<td>Carreiras, Duñabeitia, Vergara, de la Cruz-Pavia, &amp; Laka, 2010</td>
<td>object-gap relative clause preference</td>
<td></td>
</tr>
<tr>
<td>Nakh-Dagestanian</td>
<td>Avar</td>
<td>Polinsky, Gomez Gallo, Graff, &amp; Kravtchenko, 2012</td>
<td></td>
</tr>
<tr>
<td>Indo-European</td>
<td>Catalan</td>
<td>Gavarró et al., 2011</td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>Frazier, 1987</td>
<td>Mak, Vonk, &amp; Schriefers, 2002</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>Frauenfelder, Segui, &amp; Mehler, 1980</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>Schriefers, Friederici, &amp; Kuhn, 1995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mecklinger, Schriefers, Steinhauer, &amp; Friederici, 1995</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greek (Cypriot)</td>
<td>Theodorou &amp; Grohmann, 2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greek (Modern)</td>
<td>Katsika &amp; Allen, 2014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guasti, Stavrakai, &amp; Arioso, 2012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sometimes object-gap preference due to variable word order</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>Volpato &amp; Adani, 2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Di Domenico &amp; Di Matteo, 2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portuguese (European)</td>
<td>Costa, Lobo, &amp; Silva, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romanian</td>
<td>Bentea, 2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td>Polinsky, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>Betancort, Carreiras, &amp; Sturt, 2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungarian</td>
<td>McWhinney &amp; Pleh, 1988</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish</td>
<td>Kahraman et al., 2010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 15: European crosslinguistic evidence of subject-object relative clause processing asymmetry. Unless otherwise noted, findings were for subject-gap relative clause preference.
<table>
<thead>
<tr>
<th>Language</th>
<th>References</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quechuan</td>
<td>Conchucos Quechua</td>
<td>Courtney, 2006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-significant subject preference</td>
</tr>
<tr>
<td>Otomanguean</td>
<td>Ixcatec</td>
<td>Adamou, 2017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-significant subject preference</td>
</tr>
<tr>
<td>Mayan</td>
<td>Ch’ol</td>
<td>Clemens et al., 2015</td>
</tr>
<tr>
<td></td>
<td>Q’anjob’al</td>
<td>Clemens et al., 2015</td>
</tr>
<tr>
<td>Austronesian</td>
<td>Chamorro</td>
<td>Borja, Chung &amp; Wagers 2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-nominal RC subject preference; pre-nominal RC subject preference subject to dialectal variation</td>
</tr>
<tr>
<td>Indo-European</td>
<td>Portuguese (Brazilian)</td>
<td>Gouvea, 2003</td>
</tr>
<tr>
<td>Afro-asian</td>
<td>Arabic (Palestinian)</td>
<td>Botwinik, Bshara, &amp; Armon-Lotem, 2015</td>
</tr>
<tr>
<td></td>
<td>Hebrew</td>
<td>Friedmann &amp; Novogrodsky, 2004; Armon, 2005</td>
</tr>
<tr>
<td>Sino-tibetan</td>
<td>Cantonese</td>
<td>Pozniak, Huang &amp; Hemforth, 2017</td>
</tr>
<tr>
<td></td>
<td>Mandarin</td>
<td>Hu, Gavarró, Vernice, &amp; Guasti 2016; Hsiao &amp; Gibson; 2003; Lin &amp; Bever, 2006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>disagreement, but most literature points to subject-gap relative clause preference</td>
</tr>
</tbody>
</table>

Table 16: Non-European crosslinguistic evidence of subject-object relative clause processing asymmetry. Unless otherwise noted, findings were for subject-gap relative clause preference.

The subject-gap relative clause preference in languages with post-nominal relative clauses has been established in the Uralic (Hungarian: McWhinney & Pleh, 1988), Afro-asian (Hebrew: Friedmann & Novogrodsky, 2004; Armon, 2005; Palestinian Arabic: Botwinik, Bshara, & Armon-Lotem, 2015), and Sino-tibetan (Cantonese: Pozniak, Huang & Hemforth, 2017) language families and Turkish (Kahraman et al., 2010). In Japanese (Miyamoto & Nakamura, 2002; Ueno & Garnsey 2008) and Korean (Kwon, Polinsky, & Kluender, 2006; Kwon, Lee, Gordon, Kluender, & Polinsky, 2010; Yun, Whitman, & Hale, 2010), languages with post-nominal relative clauses, the subject-gap preference has also been observed. In Chamorro, which has pre- and post-nominal relative clauses, a subject preference in interpreting
ambiguous relative clauses was observed for post-nominal relative clauses, whereas for pre-nominal relative clauses, the subject preference was only clearly observed in one dialect (Borja, Chung & Wagers 2016).

In ergative the Nakh-Dagestanian language Avar, an intransitive absolutive subject preference was observed compared to transitive absolutive objects and transitive ergative subjects in a reading-time study, however, no difference was preference was observed between transitive absolutive objects and transitive ergative subjects (Polinsky et al. 2012). Stronger evidence of the subject preference in ergative languages comes from the studies in two different branches of the Mayan language family, Ch’ol and Q’anjob’al, where a subject preference in reading time and accuracy for transitive ergative subjects compared to transitive absolutive objects was found (Clemens et al. 2015).

Evidence of a subject preference was also found in two understudied indigenous American languages, Conchucos Quechua and Ixcatec, from the Quechuan and Otomanguean language families, however, the results from the study were not significant. In Conchucos Quechua speaking children and adults, a non-significant subject preference was observed in accuracy (Courtney, 2006). Similarly, non-significant subject preference in reading time was found in Ixcatec, however, the majority of ambiguous relative clauses were interpreted as subject-relatives in a picture-selection task (Adamou, 2017).

The plurality of studies (44%, n=12) of the subject-object relative clause processing asymmetry have conducted in Indo-European languages originally from Europe in the Romance, Germanic, Slavic, and Hellenic branches. Among the Indo-European Romance languages, a subject-gap preference has been found in Spanish (Betancort, Carreiras, & Sturt, 2009), Brazilian Portuguese (Gouveia, 2003), European Portuguese (Costa, Lobo, & Silva, 2011), Catalan (Gavarró et al., 2011), Romanian (Bențea, 2012), French (Frauenfelder, Segui, & Mehler, 1980), and Italian (Volpato & Adani, 2009; Di Domenico & Di Matteo, 2010). The subject-gap preference has also been observed in Dutch (Frazier, 1987; Mak, Vonk, & Schriefers, 2002), German (Schriefers, Friederici, & Kuhn, 1995; Mecklinger, Schriefers, Steinhauer & Friederici, 1995), Russian (Polinsky, 2011), and two varieties of Greek: Modern Greek (Guasti, Stavrakai, & Arioso, 2012), and Cypriot Greek (Theodorou & Grohmann, 2012). However, as Modern Greek has variable word order, the subject-object processing asymmetry can surface as either an object-gap or subject-gap preference (Katsika & Allen, 2014).
Results for Mandarin have been mixed, with Hu, Gavarró, Vernice, and Guasti (2016) supporting the subject-gap relative clause preference, and Hsiao and Gibson (2003) finding instead a preference for object-gap relative clauses. Subsequent work, such as Lin and Bever (2006) has raised issues with the methodology used in Hsiao and Gibson (2003); nevertheless, there seems to be agreement that an asymmetry is present though there is not agreement as to whether the asymmetry favors subject-gap or object-gap structures in Mandarin. Basque, an ergative, head-final language with pre-nominal relative clauses has been documented as also having a preference for object-gap relative clauses (Carreiras, Duñabeitia, Vergara, de la Cruz-Pavía, & Laka, 2010).

2.2.4 Sources of asymmetry

A number of domain-general and domain-specific sources of the subject-object relative clause processing asymmetry have been proposed. Evidence has also been found for domain-general sources of the processing asymmetry, such as additional burdens placed on working memory and statistical expectations, as well as domain-specific sources of the processing asymmetry such as syntactic reanalysis and syntactic-thematic mismatch. These proposals are not entirely mutually exclusive and provide further evidence that the subject-object relative clause processing asymmetry is an ideal linguistic phenomenon to study heritage speaker language processing given the number of factors that contribute to the asymmetry.

An asymmetry in working memory demands is one explanation. There is a greater processing cost when the object-relative is held in mind in comparison with the subject-relative. Specifically, in a subject-gap relative clause, the gap is close to the base position, whereas in an object-gap relative clause, the gap is much further from the base position. This creates a greater working memory load because the relativized noun phrase is stored, unattached, using memory resources for a longer time (Ford, 1983; Frazier & Fodor, 1978; Gibson, 1998; MacWhinney, 1987; Wanner & Maratsos, 1978).

An additional possible cause of the asymmetry is the role of frequency and expectation. The preference for subject-gap relative clauses is modulated by the animacy of the relativized noun phrase (Mak, Vonk, & Schriefers, 2002, 2006) and the semantic plausibility of the construction (Mecklinger, Schriefers, Steinhauer, & Friederici, 1995). Additionally, subject-gap relative clauses have been found to
be more common than object-gap relative clauses in corpus studies (Reali & Christiansen, 2007; Kidd, 2011). According to an emergentist or statistical-based approach these should be easier to process (Reali, & Christiansen, 2007a, 2007b; Roland, Mauner, O'Meara, & Yun, 2012).

The active filler hypothesis proposed by Clifton and Frazier (1989, et seq.) implicates the parser in the processing asymmetry. The active filler hypothesis argues that the parser actively searches for a gap to associate the filler, in this case the relativized noun phrase that is held in working memory. In subject-gap relative clauses, the first parse that associates the filler to the gap after the complementizer is valid. However, in object-gap relative clauses, the association of the filler immediately after the complementizer is incorrect when the parser encounters the noun phrase in the subject position of the subordinate clause. This failed parse and the required reanalysis of the syntactic structure of the subordinate clause incur additional processing costs for object-gap relative clauses.

Another possible contributor to the asymmetry is the mismatch between the syntactic or thematic roles of the relativized noun phrase in object-gap relatives clauses. In syntactic-based accounts, the emphasis is on the difficulty caused by a discrepancy in the roles of the head noun (Sheldon, 1974). In object-subject and subject-object relative clauses, the head noun undergoes a role reversal. In a subject-object relative clause, the subject-agent of the matrix clause becomes the patient-object in the subordinate clause, and in an object-subject relative clause, the object-patient of the matrix clause becomes the subject-agent in the subordinate clause. In each case, a processing cost is incurred because the thematic/syntactic role of the noun phrase must be reversed. However, for subject-subject relative clauses, the relativized noun phase is the subject-agent in both clauses, and in object-object relative clauses the relativized noun phrase, the patient-object of the matrix clause, is also the patient-object in the subordinate clause (Frazier & Fodor, 1978; Frazier & Clifton, 1996; Traxler, Pickering, & Clifton, 1998; Pickering, Traxler, & Crocker, 2000).

3. Previous studies of relative clause asymmetry in heritage speakers

Only a handful of studies have examined the relative clause asymmetry in heritage speakers (Korean: Lee-Ellis, 2011; Lee, 2013; Mandarin: Jia & Paradis, 2018; Russian: Polinsky, 2008, 2011). Lee (2013) found that evidence of a subject-gap preference in a picture-selection task in both early simultaneous and
early sequential adult Korean Heritage speakers raised in the anglophone US. However, the higher accuracy on subject-relatives in the early sequential heritage speaker group did not reach significance. Lee-Ellis (2011) found mixed evidence of the subject-gap preference in adult Korean heritage speakers raised in anglophone US. In an elicited production task, the early sequential bilinguals exhibited one of three patterns: subject-gap preference, object-gap preference, and no asymmetry. Korean heritage speakers with the highest Korean proficiency scores showed either a subject-gap preference (17.65%; n=3) or no asymmetry (52.94%; n=9). Polinsky (2008, 2011) found that adult Russian heritage speakers in the anglophone US had a subject-gap preference, as they performed at chance on object-relatives, whereas monolingual adults performed at ceiling on a picture-selection task. Heritage speaker children are comparable to monolingual children in comprehending sentences with relative clauses (Polinsky, 2006, 2011; Jia & Paradis, 2018). Evidence from the limited studies of heritage speaker relative clause comprehension have had mixed results: no subject-gap preference if participant performance is at ceiling, however, if performance is not at ceiling, there seems to be a subject-gap preference.

4. Empirical study of comprehension of relative clauses

The findings of Polinsky (2008, 2011), Lee-Ellis (2011), and Lee (2013) strongly suggest that: 1) the subject-object relative clause processing asymmetry is measurable with picture-matching comprehension tasks, 2) heritage speakers will be less accurate on object-relatives that subject-relatives, and 3) non-heritage speakers will perform at ceiling. We test these predictions using a picture-matching task. Additionally, before investigating heritage language processing of relative clauses using ERP and pupillometry, we need to establish that the findings are attributable to differences in processing and not comprehension. We administered and analyzed a picture-pointing task with aurally-presented Spanish relative clauses. Within and between group differences for heritage speaker and late bilingual accuracy on subject- and object-relatives are modeled using logistic mixed effects models. In the present study, we examine the likelihood of responding correctly to an item, rather than averaging a fundamentally dichotomous variable as percent accuracy, as was done in previous studies (Polinsky, 2008, 2011; Lee-Ellis, 2011; Lee, 2013).
4.1 Methodology

The RISLUS Multilingual Syntax Test (Klein & Martohardjono, 2009) was administered to heritage speaker and late bilingual participants of the ERP and pupillometry studies reported in chapters 3 and 4. The test measures auditory comprehension of complex sentences in Spanish. The test includes 9 conditions: control (subject, object), temporal adverbials (natural order, reverse order), coordination (subject, object, verb, IP), and relative clause (object-subject, subject-object, subject-subject) sentences. The present study examines only the results from the subject-subject and subject-object relative clause items.

4.1.1 Participants

Forty-seven Spanish-English bilingual adults living in the New York City participated in the study at the CUNY Graduate Center in midtown Manhattan, NY and were compensated financially for their participation. The study protocol was approved by the CUNY institutional review board and written consent was obtained.

Native Spanish-speaking participants were categorized as either heritage Spanish speakers \(n=35\) or late bilinguals \(n=12\) based on criteria commonly used in heritage speaker studies (Benmamoun, Montrul, & Polinsky, 2013) and the pre-determined inclusion criteria. Nearly two-thirds of the heritage speakers were born in the Anglophone US \(65.71\%, n=23\) and the rest moved to the Anglophone US before age 8 \(34.29\%, n=12, M=3.38, SD=2.47\). Heritage speakers were raised speaking primarily Spanish until at least age 10 by Spanish-speaking immigrant parents originally from a Spanish-dominant country/region. Late bilinguals were born in a Spanish-dominant country/region and moved to the Anglophone US at the age of 17 or older \(M=26.80, SD=4.76\).

4.1.2 Stimuli

The auditory stimuli were prescriptively grammatical complex Spanish sentences. There were 4 items per condition (subject-subject relative clause items in (23) and subject-object relative clause items in (24)) and 36 fillers (total items \(n=44\)). The items’ matrix clauses were intransitive, and the subordinate clauses were intransitive. Actors in the items were all anthropomorphic animals with masculine gender in Spanish
to avoid the confound of plausibility. The subject- and object-relative clause items included were all subject-embedded.

**Subject-relatives**

(23) *El perro, que ___ abraza al mono, duerme*  
*The dog, that ___ hugs the monkey, sleeps.*

**Object-relatives**

(24) *El mono, que el perro golpea ___ , llora*  
*The monkey, that the dog hit ___ , cries.*

Stimuli were recorded using SoundForge sampling at 44.1kHz in a sound-proof booth by a female L1 Spanish late bilingual speaker from Uruguay produced as natural running speech with neutral prosody and. All stimuli were normalized, amplified to an average loudness of -26.00 dB, and the noise filtered out using Audacity® 2.0.3 (Audacity Team, 2014), then exported as WAV files.

### 4.1.3 Protocol

The picture-selection task was administered 10-14 days after the ERP or pupillometry experiment, the primary task. Stimuli were presented over external speakers with an array of three images including two plausible distractor images in E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) as in Figure 8 (the array presented with the example item in (24)). Participants were instructed to respond to the sentences as quickly as possible by pressing a button on a serial response box. There was a 1000ms inter-stimulus interval. Instructions and 5 guided practice items with feedback wereaurally presented in Spanish.

![Figure 8: Spanish RMST object-relative clause item picture array.](image-url)
4.1.4 Data analysis

Participant-average accuracy was calculated in R to explore trends in the data. A logistic mixed-effect model was fit to the dichotomous participant item-level response data (0=incorrect, 1=correct) using the lme4::glmer() function (Bates et al., 2015). The fixed effects were condition (subject-relatives, object-relatives), bilingual group (late bilinguals, heritage speaker bilinguals), and the interaction of condition and bilingual group. By-participant and by-item random intercepts were included as random effects.

A backwards elimination procedure was used to determine the most parsimonious model. First, a maximal model with condition, group, and the interaction of condition and group. Experimental predictor variables were removed from the model in a step-wise fashion. At every step, the variable with the highest $p$-value derived from the $z$-score of the fixed effects was removed. To determine if a given variable should have been retained, models with and without the variable were compared to check if the fit of the model without the variable was significantly degraded. Models were compared with a likelihood ratio test. The process was completed upon reaching a model from which no other experimental predictor variables could be removed without significantly degrading the fit. The final models contained only significant predictor variables of accuracy.

![Figure 9](image)

**Figure 9:** Spanish RMST relative clause participant-average mean accuracy (proportion correct) across condition and bilingual group with 95% confidence interval error bars.
4.2 Results

Mean participant-average accuracy, the proportion of correct responses by participant, is plotted in Figure 9 and in Table 17. The results of proportion of correct responses by condition is interpreted below, but proportion differences by group or by condition are not tested with t-tests or ANOVAs as is common practice. These statistical tests are inappropriate for proportion data that is at the upper bound, that is close to 100% or 1.00 accuracy. When accuracy is very high, referred to “at ceiling”, this precludes the use of statistical tests which assume a Gaussian distribution as the accuracy data is very skewed.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Late</td>
<td>Heritage</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bilinguals</td>
<td>Speakers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject-Relatives</td>
<td>.92</td>
<td>.90</td>
<td>.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.22)</td>
<td>(.19)</td>
<td>(.20)</td>
<td></td>
</tr>
<tr>
<td>Object-Relatives</td>
<td>.96</td>
<td>.86</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.10)</td>
<td>(.24)</td>
<td>(.21)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.94</td>
<td>.88</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.17)</td>
<td>(.22)</td>
<td>(.21)</td>
<td></td>
</tr>
</tbody>
</table>

Table 17: Mean participant-average accuracy (proportion correct) by bilingual group and by condition. Standard deviations in parentheses.

Heritage speakers ($M=.88$, $SD=.22$) were less accurate on relative clause items than late bilinguals ($M=.94$, $SD=.17$). Across group, participants were less accurate on object-relatives ($M=.88$, $SD=.21$) than subject-relatives ($M=.90$, $SD=.20$). Across participants and across items, participants were at ceiling on relative clause items, $M=.89$, $SD=.21$. There were group differences in the pattern of responses by group. Heritage speakers were less accurate on object-relatives ($M=.86$, $SD=.24$) than subject-relatives ($M=.90$, $SD=.19$), as would be expected if a subject-object processing asymmetry led to differences in comprehension. However, this pattern was reversed in late bilinguals. Late bilinguals were less accurate on subject-relatives ($M=.92$, $SD=.22$) than object-relatives ($M=.96$, $SD=.10$), however, this is likely due to participants being at ceiling. Participant responses are better explored by looking at individual item responses using a generalized linear model, reported below.

No predictor variables were significant in predicting accuracy for the relative clause RMST items, including models with only one predictor variable (bilingual group only: $\chi^2(3)=1.79$, $p=.18$; relative clause
type only: $\chi^2(1)=0.31$, $p=.58$). The full model was not significant, $\chi^2(3)=3.84$, $p=.28$. The main effect of group was not significant ($B=-0.32$, $SE(B)=0.83$, $z=-0.38$, $p=.70$). The estimates from the non-significant maximal model are reported in Table 18. While heritage speakers were less likely to respond accurately than late bilinguals, this difference was not significant. The main effect of relative clause type was not significant ($B=0.83$, $SE(B)=1.01$, $z=0.83$, $p=.41$). All participants regardless of bilingual group were similarly accurate on subject-relative clause and object-relative clause items. The interaction of group and relative clause type was not significant ($B=-1.33$, $SE(B)=1.04$, $z=-1.28$, $p=.20$). Only the intercept was significant ($B=3.21$, $SE(B)=0.83$, $z=3.85$, $p<.001$).

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.21</td>
<td>0.83</td>
<td>3.85</td>
</tr>
<tr>
<td>Group(HS)</td>
<td>-0.32</td>
<td>0.83</td>
<td>-0.38</td>
</tr>
<tr>
<td>RC(object)</td>
<td>0.83</td>
<td>1.01</td>
<td>0.83</td>
</tr>
<tr>
<td>Group:RC</td>
<td>-1.33</td>
<td>1.04</td>
<td>-1.28</td>
</tr>
</tbody>
</table>

**Formula in R:** `Accuracy ~ cond * gen + (1 | part) + (1 | item)`

**Table 18:** Non-significant maximal logistic mixed-effects model of accuracy.

### 4.3 Discussion

While there appeared to be group level differences in comprehension when examining participant-average accuracy, these trends in the data that were confounded with ceiling effects were not borne out in the logistic mixed-effects model. No subject-object relative clause asymmetry was evident and there were no significant group level differences. While this would traditionally be characterized as participants performing at ceiling making common statistical techniques such as t-test and ANOVAs unusable, we modeled the binary outcome variable of accuracy more appropriately. We can instead characterize their behavior in terms of likelihood. Heritage speakers, as well as late bilinguals, were not more likely to be accurate on subject-relatives than object-relatives. Additionally, heritage speakers were not less likely to be accurate on relative clause comprehension than late bilinguals. Contra the findings of Polinsky (2008, 2011), Lee-Ellis (2011), and Lee (2013), we conclude that adult heritage speakers, being fluent speakers of Spanish, are able to comprehend both subject- and object-relative clauses equally well. Therefore,
more sensitive measures of processing are needed to explore the subject-object relative clause processing asymmetry in fluent speakers.

5. Research questions & hypotheses

In the investigation of heritage speaker processing using appropriate measures (implicit psychophysiological measures: event-related potentials (ERP), pupillometry), an appropriate comparison group (late bilinguals dominant in the HL living and working in a ML society), and an appropriate linguistic phenomenon (the well-established, cross-linguistically robust, early acquired subject-object relative clause processing asymmetry that is similar in the HL and ML), we ask the following two research questions:

1. Do late bilinguals show psychophysiological responses related to greater processing difficulty for object-relatives than subject-relatives in their first-learnt language in:
   a. an event-related potentials experiment, and
   b. a pupillometry experiment?

2. Do heritage speakers process object-relatives and subject-relatives differently when compared to late bilinguals in their first-learnt language in:
   a. an event-related potentials experiment, and
   b. a pupillometry experiment?

While ERP and pupillometry studies of the subject-object relative clause processing asymmetry have principally been conducted on monolingual participants (see chapters 4 and 5, respectively, for a review of the relative clause processing literature), we nevertheless expect that late bilinguals, being dominant in the L1 and late acquirers of the L2 and serving as the baseline/comparison group, will show greater ERP amplitude and task-evoked pupillary responses to object-relatives as compared to subject-relatives.

Our second research question is whether heritage speaker bilinguals process relative clauses in their HL similarly to late bilinguals. This will be assessed by comparing ERP responses and task-evoked pupillary responses by bilingual group. Heritage speaker bilinguals’ command, use, and low confidence in their first-learnt language, Spanish, has been discussed in chapter 2 and their underperformance on
metalinguistic tasks has been discussed and shown in chapter 2. These effects might impact ERP responses and task-evoked pupillary responses. On the other hand, relative clauses in the first learnt language, Spanish, are structurally similar to relative clauses in English, and the subject-object relative clause processing asymmetry is robust and early acquired. Likewise, evidence of high levels of relative clause comprehension suggest that a subject-object relative clause processing asymmetry will be observed. Therefore, we hypothesize that heritage speakers will process relative clauses similar to late bilinguals. If heritage speaker bilinguals process relative clauses in a similar fashion to late bilinguals, we would expect greater ERP amplitude and task-evoked pupillary responses to object-relatives as compared to subject-relatives.

However, while heritage speaker bilinguals are indeed fluent speakers of Spanish, they are characteristically dominant in their second-learnt language, English. Therefore, if we do observe differences in ERP amplitude and task-evoked pupillary responses between the bilingual groups, this might indicate that language processing is affected by factors contributing to dominance, such as patterns of language use. If heritage speaker bilinguals process relative clauses in a significantly different way than late bilinguals or a processing asymmetry is not observed, this should be reflected in their ERP amplitude and task-evoked pupillary responses.

6. Summary
This chapter presented the rationale for the ERP and pupillometry experiments that will be presented in the following chapters. Motivation for the linguistic phenomenon explored in the following experiments was presented and reviewed by providing evidence of the robustness of the phenomenon making it an ideal candidate for investigation. Previous studies of heritage speaker comprehension of subject- and object-relative clauses showed that heritage speakers exhibit an asymmetry measurable as poor comprehension of object-relatives. These claims are refuted by a study of heritage speakers’ high performance on subject-relatives as well as object-relatives. The major research questions addressed in the following chapters and our hypotheses were outlined.
CHAPTER IV:
EVENT-RELATED POTENTIALS

The present chapter will report the results of a study of relative clause processing in heritage speakers and late bilinguals using neuroelectrical measures. Event-related potential (ERP) studies utilize changes in voltage measures at the scalp using electroencephalic recordings that are timed to some task or event. Section one of the chapter reviews studies utilizing ERPs in the study of the subject-object relative clause processing asymmetry, and multilingual language processing. Section two presents the research questions and hypotheses. Section three explains the methodology used in the present study. Section four explores the unprocessed ERP data recorded. Section five reviews the data analyses completed and section six presents the results of the study. Section seven discusses the findings from section six. Section eight presents the conclusions.

1. ERP & language processing

1.1 Language ERP components

A number of ERP components, or deflections, have been observed in response to linguistic stimuli and taken to reflect aspects of online language processing. These components are characterized in terms of their polarity (positive or negative), latency (when the deflection peaks or reaches the local maximum), and spatial distribution (at which electrodes on the scalp the deflection is greatest in amplitude). ERP components are observed by comparing the waveforms from one stimuli type to another stimuli type. The majority of ERP studies on language processing have looked at the ERP components elicited by morphosyntactically or semantically illicit structures relative to the licit structures (for a comprehensive review of ERP and language processing see Kutas et al., 2007). Some of the most commonly found ERP components that are indicative of different aspects of language processing are the N400, late anterior negativity, and P600. The late anterior negativity (LAN), first discussed by Friederici, Pfeifer, and Hahne (1993), is a negative-going deflection observed 300-500 ms post-critical event and is maximal over left anterior electrode sites. The LAN has been argued to reflect syntactic violation (and not semantic) processing (for a thorough review of the LAN literature see Higby, 2017). The N400, first talked about by
Kutas and Hillyard (1980), is similar to the LAN in terms of latency and polarity but differs in topography. It is a negative-going deflection observed 300-500 ms post-critical event and is maximal over centroparietal electrode sites. The N400 has been argued to reflect processing of semantic anomalies and semantic integration of a noun phrase (for a thorough review of the N400 literature see Higby, 2017 and Ledwidge, 2017). The P600 is a slow, positive-going deflection observed 500-700 ms post-critical event and is maximal over centroparietal electrode sites. The P600, first talked about by Osterhout and Holcomb (1992), has been argued to reflect the processing of syntactic violations (for a thorough review of the P600 literature see Higby, 2017 and Ledwidge, 2017). In addition to reflecting the processing of syntactic violations, the P600 has also been argued to reflect difficulty in syntactic integration (Coulson, King, & Kutas, 1998; Kaan, 2007), syntactic processing (Gouvea et al., 2010; Phillips, Kazanina, & Abada, 2005), or the processing of a non-preferred event (Friederici, 1995; Carreiras et al., 2004).

1.2 ERP components & relative clause processing

In the processing of relative clauses (and other filler-gap constructions) a noun phrase is held in working memory until it can be associated to a gap. Unlike the majority of ERP studies in language, the comparison of subject- versus object-relative clauses does not contrast the two conditions in terms of a contrast in grammaticality or anomaly, but rather the distance between the relativized noun and the gap.

Previous ERP studies of subject-object relative clause processing asymmetry in Basque (Carreiras et al., 2010), Chinese (Yang, Perfetti, & Liu, 2010), English (King & Kutas, 1995; Müller, King & Kutas, 1997; Wang et al., 2015), German (Mecklinger et al., 1995), Japanese (Ueno & Garnsey, 2007), and Korean (Kwon et al., 2013) have found N400, LAN, and/or P600 ERP language components similar to ERP studies exploring ungrammaticality or anomaly. Researchers argue that the ERP components observed in response to relative clause processing reflect the working memory demands for storage and maintenance, filler-gap association, and syntactic reanalysis. However, there is no consensus as to which ERP language components reflect which cognitive process. Both positivities and negativities with differing spatial distributions have been found and have been taken to reflect the subject-object relative clause processing asymmetry. The following is a brief overview of this literature.
Some researchers (King & Kutas, 1995; Mecklinger et al., 1995; Müller, King, & Kutas, 1997; Ueno & Garnsey, 2007; Yang, Perfetti, & Liu, 2010; Kwon et al., 2013; Wang et al., 2015) found that object-relatives elicited a broadly distributed negativity that is strongest over anterior electrode sites. This negativity, which is similar to a LAN or N400, is taken to reflect successful integration of the subject filler (King & Kutas, 1995; Mecklinger, 1995), higher working memory demands that result from the association (Ueno & Garnsey, 2007; Wang et al., 2015), or higher working memory demands resulting from storage of the filler (Yang, Perfetti, & Liu, 2010). The negative-going deflection is typically sustained and was observed in the 300-500 ms time window post-violation onset. The polarity and latency of these negativities is consistent with a LAN or N400, although the spatial distribution of these negativities is broader and less localized than proto-typical LANs or N400s in that they are not restricted, or even maximal, over left anterior or centroparietal electrode sites, respectively.

Other findings include: sustained positivities in addition to the aforementioned negativities for object-relatives compared to subject-relatives (Mecklinger et al., 1995; Ueno & Garnsey, 2007; Yang, Perfetti, & Liu, 2010; Wang et al., 2015); positivities only elicited by the more difficult to process relative clause, which is the subject-relative in the case of Basque (Carreiras et al., 2010). These positivities were also taken to also reflect syntactic integration costs (Ueno & Garnsey, 2007; Wang et al., 2015), syntactic reanalysis (Mecklinger et al., 1995), or working memory demands for storage of the filler (Yang, Perfetti, & Liu, 2010). The positive-going deflection found was a late, slow wave consistent in polarity and latency with the P600.

In other ERP studies of filler-gap dependencies, researchers found LAN effects in questions (Kluender & Kutas, 1993a; Feibach et al., 2001, 2002; Epstein et al., 2013) that are argued to reflect filler maintenance in working memory. Researchers have also found ERP components which index the integration of a filler at the gap site in questions: a LAN effect (Kluender & Kutas, 1993b; Weckerly & Kutas, 1999), a very slow frontal positive wave (Kluender & Kutas, 1993a), and a P600 effect (Fiebach et al. 2001, 2002). These studies further support the finding of a LAN-like effect for object-relative clauses compared to subject-relatives. However, if integration of the filler in subject-relatives elicits a P600 effect, then this might mask a P600 effect elicited by syntactic reanalysis or filler memory maintenance in object-relatives.
While previous ERP research has found LAN- and N400-like ERP components as well as P600-like ERP components that reflect the subject-object relative clause processing asymmetry, these findings are not necessarily contradictory. Previous research on ERP components associated with morphosyntactic processing have found biphasic N400-P600 responses, responses where an N400 effect is followed by a P600 effect (e.g., Kolk et al., 2003; Hoeks, Stowe, & Doedens, 2004; Meulman et al., 2014), and biphasic LAN-P600 responses, where a LAN is followed by a P600 effect (e.g., Tanner & van Hell, 2014; Regel, Meyer, & Gunter, 2014). These findings are in fact quite common (see van Petten & Luka, 2012 for a review of 45 ERP studies). A biphasic effect might not have been observed in the previous relative clause ERP studies due to task-specific effects (Schact et al., 2014) or the fact that spatiotemporal overlap of the N400 and P600 components can attenuate the P600 response (Brouwer et al., 2016; Brouwer & Crocker, 2017).

1.3 ERP components & multilingual processing

In ERP studies where different types of bilinguals, such as early and late adult learners, are compared to monolinguals, a number of differences are observed: lack of monolingual ERP components (see Kotz 2009 for a review of ERP and fMRI studies), delayed latency (see Moreno et al., 2008 for a review of N400 latency delays in bilinguals), decreased amplitude, and spatial distribution differences. These effects are impacted by various factors including age of acquisition, L2 proficiency, and similarity between the L1 and L2 (Van Hell et al., 2010). These differences have also been interpreted as distinct processing strategies by bilinguals that then change as the bilingual becomes more proficient (Osterhout et al., 2008).

The difference between adult late learners and monolinguals is not entirely unexpected given the proficiency differences between the two groups. Late exposure and lower vocabulary size were associated with longer latency of the N400 and attenuated amplitude of the P600 in Spanish-English bilinguals (Moreno & Kutas, 2005). Felser and Clahsen (2009) argue that these differences, which are similar to patterns seen in children relative to adults, reflect inefficient processing. In addition to differences in latency (Rossi et al., 2006) and attenuated amplitude, late L2 learners have shown altogether different ERP component patterns (Mueller et al., 2005; Rossi et al., 2006; Guo et al., 2009).
possibly reflecting different processing strategies entirely that are modulated by proficiency (Ojima et al., 2005) and age of arrival (Nichols & Joanisse, 2017). Few studies control for both and proficiency and age of arrival are often confounded. In addition to proficiency and age of arrival, motivation to speak like a native speaker impacts ERP component patterns in proficient L2 learners (Tanner et al., 2014). In a meta-analysis of 41 ERP papers, Caffarra and colleagues (2015) looked at the effect of age of arrival, L1-L2 similarity, proficiency, and immersion on the detection of ELAN, LAN, N400, and P600 effects. They found that LAN effects were significantly more likely when the L2 learner spent more than 5 years in an L2-speaking country. They also observed that N400 responses to L2 syntactic violations in bilinguals were more likely when the bilinguals had an early age of arrival. Furthermore, P600 effects were significantly more likely when the bilingual participant was proficient in the L2.

In spite of these differences, bilinguals are not oblivious to grammatical and syntactic rule violations. ERP components have been observed indicating that learners are automatically processing sentences syntactically (Isel, 2007). While at first, different processing strategies are used, L2 learners eventually process language similarly to native speakers (Tanner et al., 2009, 2013), even though some differences may remain in the processing of complex syntax (Clahsen & Felser, 2006). High proficiency bilinguals show the same ERP component patterns as monolinguals (Hanna et al., 2016) and only some differences in amplitude of the ERP components may remain (Rossi et al., 2006).

Bilinguals who acquired the L2 earlier were more similar to monolinguals native speakers. For individuals who learned their L2 before age 11, no differences in latency were observed, whereas those who learned after age 11 did have latency delays which was argued to be related to proficiency (Weber-Fox & Neville, 1996). There is conflicting evidence for the exact age of acquisition that is the cut-off for monolingual-like performance. Kotz and colleagues (2008) found that individuals who acquired the L2 at an early age (before 6 years old) generally patterned like native speaker monolinguals but still showed significant differences in latency and spatial distribution of ERP components. This suggests that factors other than age of acquisition or exposure may be at play.

The effects of proficiency and age of exposure are not limited to bilinguals on the temporal and spatial characteristics of ERP components. While not often measured in monolinguals due to the assumption that all monolingual adults have similar grammars, monolinguals may have different levels of
proficiency (e.g., monolinguals who acquired their language late that was more frequently the case for deaf sign language users born to hearing parents who were not exposed to language until school-age, monolinguals that did not have access to post-secondary education). Lower proficiency in monolinguals\(^4\) resulted in more widespread spatial and temporal distribution of components which Pakulak and Neville (2010) argue result from inefficient processing strategies. Weber-Fox and colleagues (2003) found that, in monolinguals, lower proficiency was associated with delays in latency. In addition to monolingual proficiency effects, age of exposure to the L1 has shown to impact lateralization in monolingual, American sign language using deaf adults (Neville et al., 1997; Newman et al., 2002).

Differences between bilingual and monolingual groups are attenuated by structural similarity of the two languages. L2 rules that are not in the L1, for example, do not elicit any ERP components associated with ungrammaticality until later stages of acquisition (McLaughlin et al., 2010). In contrast, when the syntactic rules between the L1 and the L2 were similar, advanced L2 learners exhibited similar ERP components to those of monolinguals (Foucart & Frenck-Mester, 2011; Zawiszewski et al., 2011). Likewise, if the grammatical rules overlap in the L1 and L2, early bilinguals pattern like monolinguals, but pattern like late L2 learners when the grammatical rules between the L1 and L2 differ (Díaz et al., 2016). However, if the languages are typologically similar and the same grammatical rule is present in both languages but realized slightly differently differences in processing are observed. For example, in Spanish and Catalan where similar but distinct morphological rules hold for irregular verbs, differences between bilinguals and monolingual ERP patterns are observed (de Diego Balaguer et al., 2005). Differences between bilinguals and monolinguals in language processing are not fixed; bilingual ERP signatures can approximate monolingual ERP signatures with high proficiency and structural similarity between the L1 and L2. Proficiency is an issue for all language users but is often not studied or measured in monolinguals due to erroneous assumptions.

In the present study, which looks at relative clause processing in heritage speakers and late bilinguals, heritage speakers are not L2 learners of Spanish. Spanish is their L1 and English was

\(^4\) Proficiency is typically only measured and discussed in multilinguals. Monolingual proficiency is almost never measured as adults are assumed to have fully mastered the only language they use. This fails to acknowledge that certain structures in a language might only be present in an academic variety or a variety controlled and shaped by a high-status community. Monolingual proficiency in the cited studies refers to performance on standardized measures of language ability such as vocabulary assessments.
acquired simultaneously or at an early age (before age 6). There is a dearth of ERP literature on language processing by heritage speakers, but given the group characteristics, we can expect them to pattern like early, high proficiency L2 learners.

2. Research question & hypotheses

Previous literature has established that ERP language components are observed in response to the subject-object relative clause asymmetry in monolingual populations. We aim to extend this research by looking at the processing of relative clauses in bilinguals: late bilinguals and heritage speakers. The ERP experimental paradigm is well-suited methodologically for heritage speaker bilingual studies as the effects of linguistic insecurity are mitigated. We ask the following two research questions:

1. Are ERP components which indicate processing difficulty greater for object-relatives than subject-relatives in late bilinguals?
2. Do heritage speakers process object-relatives and subject-relatives differently when compared to late bilinguals when hearing Spanish sentences?

The subject-object asymmetry has been observed across a number of experimental methodologies and languages. We expect that the increased processing costs associated with object-relative clauses should be observable as differences in ERP language components. We expect that object-relative clauses, being more difficult to process, should elicit an ERP component of greater magnitude for object-relatives than subject-relatives. Both positive and negative polarity ERP components have been observed indexing greater processing difficulty. Therefore, we do not make any specific predictions regarding the polarity characteristics of the ERP component. Additionally, the present study was not designed to explore whether processing difficulty arises from integration difficulty, non-preference, and/or syntactic ambiguity resolution. These different processes have been argued to impact the distribution of the ERP components. Likewise, great variability in the spatial and temporal distribution of ERP components has been observed in most populations, including bilinguals and monolinguals. Therefore, we do not make any specific predictions regarding the strict spatial characteristics of the ERP component we might identify. That is not to say that we are looking for any ERP effect with any latency, spatial, and polarity
characteristics, but rather that we will aim to identify the presence of ERP components consistent with previous literature, namely the LAN, N400, and P600, and take these to be indicative of processing difficulty.

Our second research question is whether heritage speaker bilinguals process relative clauses similarly to late bilinguals. This will be assessed by comparing ERP responses by bilingual group. Heritage speaker bilinguals’ command, use, and low confidence in their first-learnt language, Spanish, has been noted in the literature and their underperformance on metalinguistic tasks has been documented. These effects might have an impact on ERP component amplitude. On the other hand, relative clauses in the first learnt language, Spanish, are structurally similar to relative clauses in English. Therefore, assuming positive transfer from the dominant language, we might hypothesize that heritage speakers will process relative clauses similarly to late bilinguals. If heritage speaker bilinguals process relative clauses in a similar fashion as late bilinguals, we would expect ERP component responses for heritage speaker bilinguals would not to be significantly different from late bilinguals.

However, while heritage speaker bilinguals are indeed fluent speakers of Spanish, they are characteristically dominant in their second learnt language, English. Therefore, if we do observe differences in ERP component responses between the bilingual groups, this might indicate that language processing is affected by factors contributing to dominance, such as use and proficiency. If heritage speaker bilinguals process relative clauses in a significantly different way than late bilinguals, this should be reflected in the amplitude for the ERP component responses. For example, heritage speakers’ ERP component responses could be significantly smaller in amplitude than observed in late bilinguals for the same ERP component.

3. Methods

3.1 Participants

Thirty-eight Spanish-English bilingual adults living in the New York City participated in the study at the CUNY Graduate Center in midtown Manhattan, NY and were compensated financially for their participation. The electroencephalography (EEG) recordings were made in the first of two test sessions. The participants were fluent Spanish and English speakers, had normal or corrected-to-normal vision,
had normal hearing, and had no history of a neurological disorder. All participants were right-handed based on the Edinburgh Handedness Inventory (Oldfield, 1971).

Demographic characteristics and language background were elicited in a 38-item questionnaire that was administered during the recruitment process. Participants were categorized as either heritage Spanish speakers ($n=18$) or late bilinguals ($n=20$) based on criteria commonly used in heritage speaker studies (Benmamoun, Montrul, & Polinsky, 2013). Heritage speakers were born in the mainland anglophone US ($n=14$) or moved to the mainland anglophone US before age 8 ($n=4$, $M=1.52$, $SD=0.96$), were raised by Spanish-speaking immigrant parents originally from the Dominican Republic, Cuba, or Puerto Rico, and were raised speaking primarily Spanish until at least age 10. Late bilinguals were born in Latin American or Puerto Rico and moved to the anglophone US at the age of 17 or older ($M=25.63$, $SD=4.67$).

Spanish proficiency in heritage speakers, who are typically English dominant, was tested using the RISLUS Multilingual Syntax Test (RMST) (Klein & Martohardjono, 2009) described in chapter 3, section 4.1. All heritage speakers performed at ceiling (Accuracy: $M=.91$, $SD=.29$, range: .77-1.00) on the RMST, indicating high ability in the comprehension of complex sentences.

3.2 Stimuli

The aurally-presented stimuli consisted of declarative Spanish sentences with either an object-subject relative clause, as in (25) below ($n=15$) or an object-object relative clause as in (26) below ($n=15$). In (25), the filler noun *la película* ‘the film’ is the object of the matrix clause *Jorge vio la película* ‘Jorge saw the film’. This relativized noun is the subject of the subordinate clause *la película ganó el premio* ‘the film won the prize’ however, there is a gap in the subordinate subject position. In (26), the filler noun *el fuego* ‘the fire’ is the object of the matrix clause *el bombero apagó el fuego* ‘the firefighter put out the fire’. This relativized noun is the object of the subordinate clause *el niño prendió el fuego* ‘the child started the fire’.

The stimuli and data in the present study are part of a larger study reported in Martohardjono and colleagues (2016, 2017). Each target stimulus was followed by a grammatical or ungrammatical question about the declarative sentence. We only analyzed data from the declarative which was the first presentation of the relative clause sentence. The comparison point for each stimulus, where the epoch
began, is the onset of the subordinate verb (||) in declarative sentences with subject- and object- relative clauses. Both constructions are grammatical and have a similar word order in English and Spanish.

Subject-Relatives (RCOS)

\[
\text{Jorge vio la película [que ganó el premio]}
\]
\[
\text{Jorge saw the movie [that won the prize]}
\]

Object-Relatives (RCOO)

\[
\text{El bombero apagó el fuego [que el niño prendió]}
\]
\[
\text{The firefighter put out the fire [that the child started]}
\]

Stimuli were recorded in a sound proof booth using SoundForge and were produced by a female L1 Spanish late bilingual speaker from Mexico as natural running speech with neutral prosody. All stimuli were normalized, amplified to -26.00 dB, and the noise filtered out using Audacity® 2.0.3 (Audacity Team, 2014), then exported as WAV files.

3.3 Experimental design & procedures

All target stimuli in the two conditions (n=30) were grammatical and were interspersed with grammatical and ungrammatical sentences (n=590) presented in the same context sentence-target question pair format. These other sentences represent conditions reported in other studies and served as fillers for the present study. Trials from different conditions were evenly distributed over five blocks and pseudorandomized so that items from the same condition never appeared consecutively. Trials were presented using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). The ERP portion of the experiment took approximately 3 hours, including preparing the participants for the task, which consisted of five blocks lasting approximately 15 minutes each with short inter-block breaks.

Participants were seated in a padded plastic chair 70cm from a computer monitor and external speakers in a shielded IAC booth. They were instructed to minimize eye movements and eye blinks. Instructions were presented aurally in English and Spanish before a short practice session.

Participants were not asked to make any metalinguistic judgments about the target items in the ERP experiment. A fixation marker (a cross in the center of the computer monitor) was provided throughout the aural blocks. Following 40% of the auditorily-presented trials, the participants were
prompted on screen to answer a comprehension question about the sentences in the trial. The comprehension question was unrelated to the research questions in this study and served the purpose of ensuring participants’ continued attention. The auditory block resumed after participants answered the question.

3.4 Data acquisition

The EEG was recorded continuously during the experiment for 32 Ag/AgCl sintered electrodes mounted in a QuikCap, positioned according to the 10-20 International Electrode System (Jasper, 1958). Data was collected for: FP1, FP2, F7, F3, FZ, F4, F8, FT7, FC3, FCZ, FC4, FT8, T7, C3, CZ, C4, T8, TP7, CP3, CPZ, CP4, TP8, P7, P3, PZ, P4, P8, O1, OZ, O2, M1, M2 with a 1000Hz sampling rate. Electrical impedances were kept below 15kΩ, and each electrode was referenced online to the electrode placed on the nose. EEG data were amplified with a NeuroScan SynAmps² amplifier and recorded between DC–100Hz. Eye movements and blinks were monitored by electrodes placed around the eyes.

3.5 Data processing

Single-trial epochs of EEG were timelocked to the onset of the subordinate verb in target stimuli offline. Each epoch began 200ms prior to the onset of the subordinate verb and lasted for 1000ms. After recording, the data were pre-processed with NeuroScan Edit software (NeuroScan Labs, Sterling, USA). Continuous EEG recordings were visually inspected for blocked or disconnected channels. Missing data from blocked and disconnected electrodes were imputed by linear interpolation from neighboring electrode sites. The data were filtered with a digital 0.1–30Hz (FIR) bandpass filter. Ocular artifacts (i.e., blinks) were corrected using the NeuroScan Spatial SVD (independent component analysis; ICA) and Spatial Filler functions. The EEG data were then segmented into individual epochs and baseline corrected by-epoch by subtracting the mean of the 200ms pre-stimulus interval. The ERP data were last re-referenced offline to the averaged mastoids. Epochs were rejected that contained artifacts exceeding +/- 70µV from the onset of the trigger to 900ms after the trigger. Four participants (heritage speakers: n=3; late bilinguals: n=1) were excluded from further analysis as more than 40% of their epochs were lost following data pre-processing in NeuroScan.
Artifact free ERP data were recorded at 1000Hz and were then exported to R (R Core Team, 2018). All data processing and later data analysis was done in R using the {rmisc} (Hope, 2013), {reshape2} (Wickham, 2007), and {plyr} (Wickham, 2011) packages. Unless otherwise specified, plots are made using {ggplot2} package (Wickham, 2009). The epochs were grand-mean centered and downsampled to 250Hz due to oversampling. The ERP components of interest are 2-5Hz waves that would require a sampling rate of only 15Hz or so to be properly captured. In order to pick the best downsampling strategy, three different methods were compared by looking for the method that had the smallest amount of variance. The three methods are:

- **4ms bin average**: The average of 4ms bins
- **Decimation by a factor of 4**: Sampling every fourth data point
- **Smoothing then decimation by a factor of 4**: Smoothing data with 4ms moving average window, then sampling every fourth data point.

This process was completed by-item, by-participant, and by-electrode. The average variance (standard deviation) was computed by-item and by-participant. The average variance for each method was compared with paired t-tests.

The method with the least amount of variance was the 4ms bin average ($M=8.34$). This was significantly better than the decimation by a factor of 4 ($M=8.36$), $t(1804)=28.22$, $p<.001$. The third method, smoothing then decimation by a factor of 4 ($M=8.35$) was neither significantly worse than the 4ms bin average method ($t(1804)=-0.03$, $p=.98$) nor significantly better than just decimation by a factor of 4 ($t(1804)=-0.16$, $p=.87$). Therefore, ERP data for relative clauses were downsampled by using 4ms bin averages for visualization and analysis.

4. Voltage data

Grand mean waveforms by relative clause condition are plotted in Figure 10 in an array with placement of each waveform in the general area of the electrode site on the scalp. Grand mean data are smoothed with a 100ms rolling average window by-electrode for plotting. Scalp arrays were made using the {erpR::scalp()} function (Arcara & Petrova, 2014) plotting the front of the head at the top of the plot.
The effect of condition, relative processing of subject-relatives versus object-relatives, is best seen by examining the subtraction wave. To derive the subtraction wave showing relative amplitude between conditions, mean subject-relative amplitude was subtracted from mean object-relative amplitude by-participant. Mean subtraction wave amplitude smoothed with a 100ms rolling average window by-electrode over the whole 1200ms epoch is plotted in Figure 11.

**Figure 10:** Grand mean waveforms for subject-relatives (solid, blue line) and object-relatives (dashed, red line) by-electrode by-millisecond smoothed with a (100ms) filter.

**Figure 11:** Grand mean waveforms for subtraction waveforms by-electrode by millisecond smoothed with a (100ms) filter.
Visual inspection of grand mean waveforms by relative clause condition and the subtraction waveforms across bilingual group, showed that compared to subject-relatives, object-relatives elicited an anteriorly- and centrally-distributed, slow, late positivity over the left hemisphere that is consistent with the polarity and spatial characteristics of a P600; a left-lateralized, anterior, slow, early negativity that is consistent with the polarity and spatial distribution of a LAN; and a bilateral, posterior, slow, early negativity that is consistent with the polarity and spatial distribution of an N400. To further examine the latency of the waveforms, the grand mean waveforms of electrode over which the P600-, LAN-, and N400-like effects are maximal are plotted in Figures Figure 12-Figure 14.

Figure 12: Grand mean waveforms by individual left and central electrode by millisecond smoothed with a (100ms) filter.
The late slow wave, a left-lateralized positive-going deflection, is broadly distributed over anterior to posterior electrode sites (F3, Fz, FC3, FCz, C3, Cz, CP3, CPz, P3, Pz). In examining the positive-going deflection plotted in Figure 12, there appears to be two late, slow positive waves: one waveform starting around 200ms, peaking around 500ms, and maximal over anterior electrode sites (F3, Fz, and FC3) that is consistent with a frontal P600; another waveform starting around 350ms, peaking at the end of the epoch around 1000ms, and maximal over centroparietal electrode sites (CPz, CP3, Cz, Pz) that is consistent with a classic centroparietal P600. A frontal P600 with a more anterior distribution but similar latency and polarity characteristics to a classic P600 is not unexpected. Frontal or Anterior P600s have been reported as an index of syntactic ambiguity resolution (Hagoort et al., 1999; Kaan & Swaab, 2003; Ledoux et al., 2007) or syntactic integration difficulties with complex sentences (Friederici, Hahne, & Saddy, 2002).

![Figure 13: Grand mean waveforms by individual posterior electrode by millisecond smoothed with a (100ms) filter.](image)

The early, posterior, negative-going deflection is broadly distributed over posterior electrode sites (P3, Pz, P4). In examining the negative-going deflection plotted in Figure 13, there appears to be an early, slow wave that starts at the trigger and peaks between 350-600ms post-gap. While the early, slow wave is not consistent with an N400, the spatial distribution is consistent with the N400-like effect that is maximal over posterior electrode sites in King and Kutas (1995).

The early, left, anterior negative-going deflection over electrode sites (F4, F8, FC4, C4) is plotted in Figure 14. A left anterior slow wave that is spatially consistent with a LAN, however the sustained nature of the negative-going deflection is not. Early-occurring greater negative mean amplitude over left anterior electrode sites (F4, FC4) for object-relatives is consistent with a LAN.
Figure 14: *Grand mean waveforms by individual left anterior electrode by millisecond smoothed with a (100ms) filter.*

5. Data analysis

The brain’s electrical signature that is measured at the scalp is highly complex and consists of many different ERP components. To test for the differences in ERP components resulting from experimental conditions in the study, the linguistic ERP components must be isolated from the auditory, visual, and other cognitive ERP components co-occurring with the language processing ERP components. In addition to overlapping and complex componential responses, ERP studies generate an extremely large amount of data making simple statistical analysis quite difficult. In looking for an effect of experimental condition, data recorded by-millisecond over 29+ different electrode sites must be examined. The epochs in this study consisted of amplitude recordings up to 1000ms after the onset of the subordinate verb. Using a down-sampled sampling rate of 250Hz, 250 data points are generated by-item by-condition by-participant by-electrode totaling nearly 8 million data points. In order to analyze the large amount of data and to identify the linguistic ERP components in any ERP study, there are a number of data analysis steps that must be completed, as illustrated in Figure 15. The electroencephalography recordings are first preprocessed, as previously discussed in sections 3.4 and 3.5, by removing and correcting artifacts like electrical noise due to blinks and muscle movement and epoched by time-locking a section of the EEG.
signal to an event, in this case the onset of the subordinate verb in relative clause constructions. Then, to identify and isolate the latency characteristics of specific ERP components, a time window of interest is selected. A spatial region of interest which isolates a linguistic ERP component that is maximal over a number of electrode sites is then identified. After the data are limited to restrict the analysis in line with the latency and spatial characteristics of a linguistic ERP component, hypothesis or difference testing is done. ERP amplitude for the different linguistic ERP components is compared by- and across-bilingual group using statistical tests to establish the relationship between the linguistic ERP components.

![Figure 15: General ERP data analysis flowchart.](image)

Traditional ERP analysis begins with the visual inspection of voltage data identifying specific electrode sites that are consistent with pre-identified ERP components. The researcher then averages over these electric sites that are selected as the region of interest within a time window determined by the researcher and informed by previous literature. The researcher typically does this by looking for ERP waveform signatures that are consistent with previously identified ERP components in terms of polarity, latency, and scalp distribution. This process involves a large number of decision points that can inadvertently bias the results. An additional confound is present in the amplitude that is recorded at the scalp at 29 electrode sites. Amplitude at each electrode site is not independent. The dipole nature of neuroelectrical activity causes voltage recorded at electrode sites at opposite ends of the head to be inversely correlated, while voltage recorded at adjacent electrode sites are highly correlated.

Instead of selecting individual electrode sites for statistical analysis or dividing and grouping electrode sites by laterality and anteriority within researcher-defined time windows (cf. Phillips, et al., 2005), we used the ERP amplitude data recorded at all electrode sites and all timepoints. These data are used to determine the componential structure of the data by performing temporo-spatial decomposition by principal component analysis (PCA). The PCA reduces the ERP data into a smaller number of virtual ERP components with distinct latency and spatial characteristics (Spencer, Dien, & Donchin, 2001; Dien &
Hypothesis testing was then done by fitting the isolated linguistic ERP component amplitude with linear mixed-effects models.

5.1 Temporo-spatial analysis

Sequential temporo-spatial analysis identifies and isolates the language processing ERP components which are then compared for effect of condition, in this study object-relative versus subject-relative. It is impossible to control all non-experimental condition related neuroelectrical activity recorded during an experiment, let alone the additional noise in the data resulting from auditory processing, therefore this study takes a data-driven “blind-source separation” approach. The ERP data undergoes temporo-spatial decomposition by PCA in order to reduce the dimensionality of the data. This allows us to identify and isolate the underlying componential structure of the ERP responses by identifying regions of interest in terms of latency and scalp distribution for later statistical analysis (Spencer, Dien, & Donchin, 2001; Dien & Frishkoff, 2005; Hestvik & Durvasula, 2016). Simply put, a temporal PCA reduces the 250 time points to only those that cluster together to make a gradient time window. The subsequent spatial PCA reduces the 29 electrode sites to clusters of electrodes whose behavior is similar. This methodology has the benefit of reducing the number of decision points that could potentially be affected by researcher bias. It is a data-
driven approach where the complex and noisy ERP data is decomposed into its underlying componential structure, which critically may not be apparent by visual inspection of voltage data. Components that are consistent with the latency, polarity, and scalp distribution of linguistic ERP components in the literature, namely P600, N400, and/or LAN, are then identified. This process is depicted in Figure 16.

A temporal PCA with a varimax is performed on the voltage data from the ERPs for all participants and items from 0-1000ms. We intended to identify at least two temporal rotated components with latency characteristics similar to either a P600 (high factor weights over late time points greater than 500ms) or an N400/LAN (high factor weights in the 300-500ms time window). The factor weights from the temporal rotated components will be used to scale the ERP amplitude by time point. A spatial PCA with a promax rotation will then be performed on the late slow wave and then another spatial PCA is performed on the early wave peaking between 300-500ms. We will then identify any spatial components consistent with the scalp distribution of a P600 component for the late slow wave (high factor weights over anterior and/or centroparietal electrode sites), an N400 component for the early wave (high factor weights over posterior electrode sites), and a LAN component again for the early wave (high factor weights over anterior electrode sites). The temporally scaled amplitude is further scaled by the loadings from the spatial rotated components to derive amplitude by-electrode, by-time point, by-participant, by-item for each of the possible three ERP components that have been previous established to index the subject-object relative clause processing asymmetry. The mean amplitude by-participant by-item across time points and electrode site will be used for difference testing with linear mixed-effects models.

5.1.1 Temporal analysis

The temporal PCA reduces the dimensionality of this data to data points that can be grouped together, which in effect defines a gradient time window of interest similar to traditional ERP analysis. Furthermore, the temporal PCA not only groups the time points into a smaller number of gradient clusters in a limited window but represents the time course of the ERP component over the whole epoch. The time course of the temporal rotated components extracted from the temporal PCA can be examined and compared to the known latency characteristics of the P600 (a late slow wave peaking after 500ms after the trigger) the N400 and/or LAN (an early wave peaking 300-500ms after the trigger).
The temporal PCA was performed over amplitude by-time point by-electrode by-item by-participant across bilingual group and across condition with the \texttt{psych::principal()} function in R (Revelle, 2016). The stopping criterion to determine the number of rotated components that should be retained for the rotation with the varimax procedure was the Kaiser-Guttman rule (Cangelosi & Goriely, 2007). 31 temporal rotated components were retained for the temporal PCA, cumulatively accounting for 93% of the total variance in the data set. Three temporal rotated components had latency characteristics consistent with a late, slow wave and an early wave peaking 300-500ms.

The first temporal rotated component accounted for 28% of variance and had latency characteristics consistent with a P600, \textit{eigenvalue}=70.54. The factor loadings for the first temporal rotated component are plotted against time point in Figure 17. The factor loadings for the first temporal rotated component are greatest in the later half of the epoch. Each data point from the beginning of the epoch is increasingly given more weight, peaking around 890ms then decreasing slightly for the remaining 110ms. The factor weights for the first rotated component are used to temporally scale the data before a spatial PCA is run to identify the scalp distribution of the late slow wave ERP components and determine if any of the resulting ERP components are consistent with an P600.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{Fig17.png}
\caption{Factor weights by millisecond for the first rotated temporal component.}
\end{figure}

The seventh temporal rotated component accounted for 4% of variance and had latency characteristics consistent with a N400/LAN, \textit{eigenvalue}=9.17. The nineteenth temporal rotated component accounted for 3% of variance and also had latency characteristics consistent with a N400/LAN, \textit{eigenvalue}=8.71. The factor loadings for the seventh and nineteenth temporal rotated components are plotted against time point in Figure 18. The factor loadings for the seventh temporal rotated component peak sharply at 375ms. The
factor loadings for the nineteenth temporal rotated component peak sharply at 420ms. Since both of these components peak within the 300-500ms time window characteristic of the N400/LAN, both are used to temporally scale the data before a spatial PCA is run to identify the scalp distribution of the early ERP components and determine if any of the resulting ERP components are consistent with an N400 and/or LAN. Note that while there appeared to be early sustained posterior negativities in the voltage data in section 4, this was not borne out in the temporal PCA. It appears that the sustained negativities are a result of a series of fast early waves.

![Early fast wave temporal components](image)

**Figure 18:** Factor weights by millisecond for the seventh and nineteenth rotated temporal component.

5.1.2 Spatial analyses

After the temporal PCA was performed to decompose the amplitude data into constituent components by latency and reduce the temporal dimensionality of the data, a spatial PCA was performed to explore the spatial distribution of the slow wave and early wave ERP components to look for ERP components consistent with the spatial distribution characteristics in addition to the latency characteristics of a P600, N400, and/or LAN. Identification of the spatial rotated component(s) exhibiting a scalp distribution consistent with the scalp distribution of a P600, N400 or LAN was done by visual inspection of temporospatially scaled grand mean by condition scalp arrays.
For the late slow wave, a spatial PCA retaining five spatial rotated components for a promax rotation was run using the same stopping criterion and R function as the temporal PCA. The five spatial rotated components cumulatively accounted for 55% of the total variance in the temporally scaled P600 data. The first spatial rotated component accounted for 15% of variance and had a scalp distribution consistent with a frontal P600, \(eigenvalue=4.41\). In examining the temporo-spatially scaled grand mean by condition scalp arrays, plotted in Figure 22, object-relatives elicited a positive-going deflection compared to subject-relatives that was maximal over anterior electrode sites. The second spatial rotated component accounted for 11% of variance and had a scalp distribution consistent with a centroparietal P600, \(eigenvalue=3.25\). In examining the temporo-spatially scaled grand mean by condition scalp arrays, plotted in Figure 26, object-relatives elicited a positive-going deflection compared to subject-relatives that was maximal over centroparietal electrode sites.

For the early wave peaking at 420ms, a spatial PCA retaining five spatial rotated components for a promax rotation was run as above. The five spatial rotated components cumulatively accounted for 63% of the total variance in the temporally scaled N420 data. The second spatial rotated component accounted for 9% of variance and had a scalp distribution consistent with a N400, \(eigenvalue=2.85\). In examining the temporo-spatially scaled grand mean by condition scalp arrays, plotted in Figure 30, object-relatives elicited a negative-going deflection compared to subject-relatives that was maximal over posterior electrode sites. No spatial rotated component scaled the data such that an LAN peaking 420ms post-event was apparent. There was no negative-going deflection for object-relatives compared to subject-relatives over left anterior electrode sites.

For the early wave peaking at 375ms, a spatial PCA retaining five spatial rotated components for a promax rotation was run as above. The five spatial rotated components cumulatively accounted for 63% of the total variance in the temporally scaled N375 data. The third spatial rotated component accounted for 10% of variance and had a scalp distribution consistent with a N400, \(eigenvalue=2.97\). In examining the temporo-spatially scaled grand mean by condition scalp arrays, plotted in Figure 34, object-relatives elicited a negative-going deflection compared to subject-relatives that was maximal over posterior electrode sites. No spatial rotated component scaled the data such that an LAN peaking 420ms post-
event was apparent. There was no negative-going deflection for object-relatives compared to subject-relatives over left anterior electrode sites.

**Figure 19:** Results of the temporo-spatial PCA ERP data analysis flowchart.

### 5.1.3 Results

The temporo-spatial PCA identified 4 ERP components consistent with language ERP components associated with subject-object relative clause processing, as illustrated in Figure 19. The frontal P600 is a slow, late, positive going-deflection peaking around 890ms and maximal over anterior electrode sites. The centroparietal P600 is a slow, late, positive-going deflection peaking around 890ms and maximal over centroparietal electrode sites. The first N400-like is an early, negative-going deflection peaking at 420ms and maximal over posterior electrode sites. The second N400-like is an early, negative-going deflection peaking at 375ms and maximal over posterior electrode sites. To distinguish these two N400-like ERP components for the remainder of the chapter, the earlier N400-like ERP component will be referred to as the N375 and the later N400-like ERP component will be referred to as the N420.
5.2 Linear mixed-effects modeling

In order to explore the subject-object relative clause processing asymmetry in heritage speaker bilinguals and late bilinguals, linear mixed-effects models with random intercepts for participant and item across groups were fit to the ERP component amplitude using the `lme4::lmer()` function (Bates, et al., 2015) and the `nloptr` package (Johnson, n.d.) in R. ERP Components were quantified using mean temporospatially scaled amplitude over the whole epoch (0-1000ms). The fixed effects were condition (subject-relatives, object-relatives), bilingual group (late bilinguals, heritage speaker bilinguals), and the interaction of condition and bilingual group.

![Figure 20: Model selection. Model fixed effects indicated with arrows indicating likelihood ratio tests comparing fit of two models.](image)

A backwards elimination procedure was used to determine the most parsimonious model, as illustrated in Figure 20. First, a maximal model with condition, group, and the interaction of condition and group was fit to the data. Predictor variables were removed from the model in a step-wise fashion. At every step, the variable with the highest p-value was removed. Degrees of freedom in order to calculate the p-values were based on the Satterthwaite approximation from the `lmerTest::summary()` function (Kuznetsova, Brockhoff, & Christensen, 2016). To determine if a given variable should be retained,
models with and without the variable were compared to check if the fit of the model without the variable was significantly degraded. Models were compared with likelihood ratio tests. The process was completed upon reaching a model from which no other variables could be removed without significantly degrading the fit. The final models contained only significant predictor variables of the ERP component amplitude. If no variables significantly improved the fit of the model, then the null, intercept-only, model is reported.

The parameters of the most parsimonious model are interpreted, though it should be noted that, due to the inclusion of two dichotomous variables, interpreting the parameters of the model is not that straightforward. In a maximal model with the variable of condition (subject-relatives, object-relatives), bilingual group (late bilinguals, heritage speaker bilinguals), and the interaction of condition and bilingual group, the parameters can be confusing. The intercept is the estimate for the reference category of condition and bilingual group: late bilinguals, subject-relatives. The parameter for the fixed effect of condition is the estimate for late-bilinguals, object-relatives. The parameter for the fixed effect of group is the estimate for heritage speakers for subject-relatives (since the reference category for condition is subject-relatives). Lastly, the parameter for the interaction of group and condition is the estimate for the difference between what is estimated based on the parameters for the effect of group and condition alone and the predicted value. A more straightforward interpretation of the interaction effect is that it shows whether heritage speakers are treating relative clauses in a similar manner to late bilinguals. A significant interaction effect can be interpreted as heritage speakers processing relative clauses (subject-relatives versus object-relatives) differently from late bilinguals. It does not identify whether heritage speakers are processing object-relatives as significantly different from subject-relatives, the primary focus of this study.

The comparisons that are the focus of this study are within-group differences between subject-relatives and object-relatives. The significance of each parameter in the maximal model does not straightforwardly represent the comparisons that would address the main research questions of this chapter: 1) Are object-relatives harder to process than subject-relatives?, and 2) Do heritage speakers and late bilinguals process relative clauses similarly? The significance of the fixed effects in a maximal model are represented in Figure 21. The significance of the main effects represents the late bilingual subject-relative ERP component amplitude compared to the grand mean ERP component amplitude, the
late bilingual object-relative ERP component amplitude, and the heritage speaker ERP component amplitude. Heritage speaker object-relative ERP component amplitude is not compared not to ERP component amplitude of any other group-condition, but rather to the sum of the estimates of group and condition.

Figure 21: Group-condition contrasts from linear mixed-effects model and from contrasts from confidence intervals for predicted marginal means from most parsimonious mixed-effects model.

In order to test the hypotheses in the present study and compare subject-relative ERP component amplitude to object-relative ERP component amplitude within-group, the predicted values from the maximal model were used to calculate the predicted marginal means with 95% confidence intervals using the \texttt{lsmeans::lsmeans()} function in R (Lenth, 2016). The observed means and the predicted marginal means are similar, but the estimates of variance differ. The significance of the difference between object-relative and subject-relative predicted ERP component amplitude controlling for individual and item level variation can be determined. The difference between group can then be established qualitatively by comparing the direction and size of the processing asymmetry effect as indexed by the ERP component amplitude.

5.3 Across ERP component analysis

Correlations of participant-average ERP component difference amplitude are run to explore the relationship between the different ERP components found in the study using the \texttt{stats::cor.test()} function in R (R Core Team, 2017). ERP component difference amplitude was calculated by subtracting mean subject-relative amplitude from mean object-relative amplitude by participant. To explore the
strength of the relationship between ERP components, mixed-effects models with ERP component difference amplitude as a fixed effect and participant and item as random effects were fit to the ERP component difference amplitude. Data used to model ERP component was scaled by ERP component in order to compare regression coefficients across ERP components using the \texttt{lme4::lmer()} function in R (Bates, et al., 2015). Significance of the fixed effect was calculated by comparing the null, intercept-only model to the model with the single fixed effect using a likelihood ratio test. The P600 effects (centroparietal P600 component amplitude predicting frontal P600 component amplitude), biphasic effects (N375 and N420 component amplitude predicting frontal P600 component amplitude; N375 and N420 component amplitude predicting centroparietal P600 component amplitude), and N400-like effects (N375 component amplitude predicting N420 component amplitude) were compared across and within bilingual group using 95% confidence interval calculated using the \texttt{lme4::coninf()} function in R (Bates et al., 2015).

6. Results

Results of scaled amplitude are reported by ERP component: frontal P600, centroparietal P600, N420, and N375. A table and plot of condition and bilingual group means is given in each section. The most parsimonious model for each ERP component is reported below and interpreted.

6.1 Frontal P600

Object-relatives, compared to subject-relatives, elicited a slow, late wave ERP component maximal over midline anterior (Fz, FCz, Cz) and left anterior (FC3, C3) electrode sites consistent with a frontal P600. A scalp array of mean scaled amplitude across group by condition by electrode site for the frontal P600 amplitude is plotted in Figure 22.
Object-relatives elicited a late slow wave starting around 200ms and peaking around 600ms. Grand mean waveforms of scaled amplitude across group across electrode sites for the frontal P600 amplitude is plotted by condition in Figure 23. Mean scaled amplitude by- and across- group and condition is reported in Table 19 and plotted in Figure 24.
Table 19: Mean frontal P600 amplitude by bilingual group and condition. (Standard deviations in parentheses).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Late Bilinguals</th>
<th>Heritage Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-Relatives</td>
<td>-338.48</td>
<td>136.42</td>
<td>-162.78</td>
</tr>
<tr>
<td></td>
<td>(2818.31)</td>
<td>(3025.86)</td>
<td>(2917.67)</td>
</tr>
<tr>
<td>Object-Relatives</td>
<td>-94.36</td>
<td>235.07</td>
<td>50.56</td>
</tr>
<tr>
<td></td>
<td>(2919.97)</td>
<td>(3288.03)</td>
<td>(3088.02)</td>
</tr>
<tr>
<td>Total</td>
<td>-241.12</td>
<td>186.78</td>
<td>-54.86</td>
</tr>
<tr>
<td></td>
<td>(2870.55)</td>
<td>(3158.65)</td>
<td>(3005.10)</td>
</tr>
</tbody>
</table>

Across group, mean scaled amplitude for object-relatives ($M=50.56$, $SD=3088.02$) was more positive than subject-relatives ($M=-162.78$, $SD=2917.67$) indicating that object-relatives elicited a greater positive-going deflection over anterior electrode sites compared to subject-relatives. Across condition, heritage speakers ($M=186.78$, $SD=3158.65$) had more positive mean scaled amplitude than late bilinguals ($M=-241.12$, $SD=2810.55$). In looking at mean scaled amplitude, both bilingual groups appear to exhibit the subject-object processing asymmetry as indexed by a frontal P600. Late bilinguals had more positive mean scaled amplitude for object-relatives ($M=-94.36$, $SD=2919.97$) than subject-relatives ($M=-338.48$, $SD=2818.31$). Likewise, heritage speakers had more positive mean scaled amplitude for object-relatives ($M=235.07$, $SD=3288.03$) than subject-relatives ($M=136.42$, $SD=3025.86$).

To determine whether differences between condition and bilingual group were statistically significant, frontal P600 amplitude was modeled following the procedure outlined in section 5.3 and visualized in Figure 25. The most parsimonious model, shown in Table 20, retained only the predictor variable of group which had a significantly better fit than the null model, $\chi^2(1)=4.625$, $p=.039$. The maximal model, shown in Table 21, with group, condition, and the interaction was not the most parsimonious model ($\chi^2(1)=0.229$, $p=.633$) nor was it a significantly better fit of the data than the null model ($\chi^2(3)=5.557$, $p=.135$).
Figure 25: Linear mixed-effects models for frontal P600 amplitude. Significant parsimonious model in red type.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-241.1</td>
<td>134.9</td>
<td>-1.787</td>
<td>--</td>
</tr>
<tr>
<td>Group(HS)</td>
<td>427.9</td>
<td>204.5</td>
<td>2.092</td>
<td>--</td>
</tr>
</tbody>
</table>

**Formula in R:** Scaled Amplitude ~ gen + (cond | part) + (1 | item)

† p<.1 † p<.05 †† p<.01 ††† p<.001

Table 20: Most parsimonious linear mixed-effects model of frontal P600 amplitude elicited by subject- and object-relative clauses.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-338.5</td>
<td>190.9</td>
<td>-2.035</td>
<td>--</td>
</tr>
<tr>
<td>Group(HS)</td>
<td>524.9</td>
<td>290.9</td>
<td>1.804</td>
<td>--</td>
</tr>
<tr>
<td>RC(object)</td>
<td>294.1</td>
<td>269.7</td>
<td>1.091</td>
<td>--</td>
</tr>
<tr>
<td>Group:RC</td>
<td>-195.5</td>
<td>408.8</td>
<td>-0.478</td>
<td>--</td>
</tr>
</tbody>
</table>

**Formula in R:** Scaled Amplitude ~ cond * gen + (cond | part) + (1 | item)

† p<.1 † p<.05 †† p<.01 ††† p<.001

Table 21: Maximal linear mixed-effects model of frontal P600 amplitude elicited by subject- and object-relative clauses.

The variable of group significantly improved the fit of the model and was a significant predictor of frontal P600 amplitude. Heritage speakers had significantly greater frontal P600 amplitude than late bilinguals, B=427.9, SE(B)=204.5, t(--)=2.092. The confidence intervals for the most parsimonious model...
parameters, shown in Table 22, further demonstrate that heritage speakers \( (B=427.9, \ SE(B)=204.5, \ 95\% \ CI \ [27.068, \ 828.734]) \) had significantly greater frontal P600 amplitude than late bilinguals \( (B=-241.1, \ SE(B)=134.9, \ 95\% \ CI \ [-505.571, \ 23.335]) \).

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>CI_lower</th>
<th>CI_upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Late Bilingual</td>
<td>-241.1</td>
<td>134.9</td>
<td>--</td>
<td>-505.571</td>
</tr>
<tr>
<td>Heritage Speaker</td>
<td>427.9</td>
<td>204.5</td>
<td>--</td>
<td>27.068</td>
</tr>
</tbody>
</table>

**Table 22:** Confidence intervals for model parameters of frontal P600 amplitude elicited by subject- and object-relative clauses.

Although the difference in mean amplitude for object-relatives across group was more positive than for subject-relatives, the predictor variable of condition did not significantly improve the fit of the model, \( \chi^2(1)=1.063, \ p=.303 \). Neither bilingual group significantly exhibited the subject-object processing asymmetry as indexed by the frontal P600.

**Figure 26:** Scalp array of scaled amplitude of the centroparietal P600 smoothed with a (100ms) filter.

6.2 Centroparietal P600

Object-relatives, compared to subject-relatives, elicited a slow, late wave ERP component maximal over centroparietal (C3, Cz, CP3, CPz, Pz) electrode sites consistent with a centroparietal P600. A scalp array
of mean scaled amplitude across group by condition by electrode site for the centroparietal P600 amplitude is plotted in Figure 26.

Object-relatives elicited a late slow wave starting around 400ms and peaking at the end of the 1000ms epoch. Grand mean waveforms of scaled amplitude across group across electrode sites for the centroparietal P600 amplitude is plotted by condition in Figure 27. Mean scaled amplitude by- and across-group and condition is reported in Table 23 and plotted in Figure 28.

Across group, mean scaled amplitude for object-relatives ($M=153.97, SD=2440.04$) was more positive than subject-relatives ($M=-81.53, SD=2399.27$) indicating that object-relatives elicited a greater positive-going deflection over centroparietal electrode sites compared to subject-relatives. Across condition, heritage speakers ($M=98.98, SD=2535.98$) had more positive mean scaled amplitude than late bilinguals ($M=-10.01, SD=2330.79$). In looking at mean scaled amplitude, it appears that only late bilinguals exhibited the subject-object processing asymmetry as indexed by a centroparietal P600. Late bilinguals had more positive mean scaled amplitude for object-relatives ($M=316.73, SD=2344.09$) than subject-relatives ($M=-338.08, SD=2275.23$). However, heritage speakers exhibited the opposite effect, they had greater positive scaled amplitude for subject-relatives ($M=257.77, SD=2520.36$) than object-relatives ($M=-53.26, SD=2548.07$).

![Figure 27: Grand mean waveform of centroparietal P600 amplitude by condition smoothed with a (100ms) filter.](image1)

![Figure 28: Mean centroparietal P600 amplitude by bilingual group and condition with 95% confidence interval error bars.](image2)
To determine whether differences between condition and bilingual group were statistically significant, centroparietal P600 amplitude was modeled following the procedure outlined in section 5.3 and visualized in Figure 29. The maximal model, which was also the most parsimonious model, shown in Table 24, retained the predictor variables of group, condition, and the interaction had a significantly better fit than the null model ($\chi^2(3) = 11.097, p = .011$) and a simpler model without the interaction ($\chi^2(1) = 8.610, p = .003$).

The intercept of the model was significant, meaning that late bilingual amplitude for subject-relatives significantly differed from the mean across condition and across group, $B = -338.1$, $SE(B) = 153.3$, $t(788.89) = -2.205$, $p = .028$. The predictor variable of group was significant, meaning that heritage speakers had significantly higher amplitude for subject-relatives than late bilinguals did for subject-relatives, $B = 595.9$, $SE(B) = 233.7$, $t(788.89) = 2.550$, $p = .011$. The predictor variable of condition was significant, meaning that late bilinguals had significantly higher amplitude for object-relatives than subject-relatives, $B = 654.8$, $SE(B) = 216.6$, $t(788.89) = 3.023$, $p = .003$. Lastly, the interaction between bilingual group and condition was significant, meaning that the difference between subject-relatives and object-relatives was significantly different between bilingual groups, $B = -965.8$, $SE(B) = 328.4$, $t(788.89) = -2.941$, $p = .003$.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-Relatives</td>
<td>Late Bilinguals</td>
<td>257.77</td>
<td>2399.27</td>
</tr>
<tr>
<td></td>
<td>(2275.23)</td>
<td>(2520.36)</td>
<td>(2344.09)</td>
</tr>
<tr>
<td></td>
<td>Heritage Speakers</td>
<td>-53.26</td>
<td>153.97</td>
</tr>
<tr>
<td></td>
<td>(2344.09)</td>
<td>(2548.07)</td>
<td>(2440.04)</td>
</tr>
<tr>
<td>Object-Relatives</td>
<td>Total</td>
<td>10.01</td>
<td>37.43</td>
</tr>
<tr>
<td></td>
<td>(2330.79)</td>
<td>(2535.98)</td>
<td>(2421.43)</td>
</tr>
</tbody>
</table>

Table 23: Mean centroparietal P600 amplitude by bilingual group and condition. (Standard deviations in parentheses).
Table 24: Maximal linear mixed-effects model of centroparietal P600 amplitude elicited by subject- and object-relative clauses.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-338.1</td>
<td>153.3</td>
<td>788.89</td>
<td>-2.205</td>
<td>.028    *</td>
</tr>
<tr>
<td>Group(HS)</td>
<td>595.9</td>
<td>233.7</td>
<td>788.89</td>
<td>2.550</td>
<td>.011    *</td>
</tr>
<tr>
<td>RC(object)</td>
<td>654.8</td>
<td>216.6</td>
<td>788.89</td>
<td>3.023</td>
<td>.003    **</td>
</tr>
<tr>
<td>Group:RC</td>
<td>-965.8</td>
<td>328.4</td>
<td>788.89</td>
<td>-2.941</td>
<td>.003    **</td>
</tr>
</tbody>
</table>

Formula in R: Scaled Amplitude ~ cond * gen + (cond | part) + (1 | item)

† p<.1 "p<.05 ""p<.01 """p<.001

The predicted values from the maximal model were used to calculate descriptive statistics, shown in Table 25 and as outlined in section 5.3, to compare groups and explore the subject-object asymmetry by- and across-group after having controlled for individual and item level variance. Across condition, although heritage speakers ($M_p=102.259$, $SE_p=123.386$, 95% $CI_p=[-139.946, 344.463]$) did have higher mean amplitude than late bilinguals ($M_p=-10.676$, $SE_p=108.303$, 95% $CI_p=[-223.272, 201.920]$), this difference was not significant. Similarly, across group, although object-relatives ($M_p=131.737$, $SE_p=115.346$, 95% $CI_p=[-94.685, 358.158]$) elicited more positive amplitude than subject-relatives ($M_p=-40.154$, $SE_p=116.829$, 95% $CI_p=[-268.486, 189.179]$), this difference was not significant. Late bilinguals
were the only group to show the subject-object asymmetry as indexed by the centroparietal P600. For late bilinguals, object-relatives ($M_p=316.729$, $SE_p=153.008$, 95% CI$_p$ [16.378, 617.080]) elicited significantly more positive amplitude than subject-relatives ($M_p=-338.081$, $SE_p=153.319$, 95% CI$_p$ [-639.041, -37.120]). For heritage speakers, object-relatives elicited the opposite effect; subject-relatives ($M_p=257.773$, $SE_p=176.322$, 95% CI$_p$ [-88.342, 603.889]) elicited a more positive amplitude than object-relatives ($M_p=-53.256$, $SE_p=172.648$, 95% CI$_p$ [-392.160, 285.648]), although this difference was not significant.

<table>
<thead>
<tr>
<th>Group</th>
<th>Relative Clause</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>CI$_{lower}$</th>
<th>CI$_{upper}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Late Bilingual</td>
<td>Subject</td>
<td>-338.081</td>
<td>153.319</td>
<td>788.89</td>
<td>-639.041</td>
<td>-37.120</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>316.729</td>
<td>153.008</td>
<td>788.89</td>
<td>16.378</td>
<td>617.080</td>
</tr>
<tr>
<td>Heritage Speaker</td>
<td>Subject</td>
<td>257.773</td>
<td>176.322</td>
<td>788.89</td>
<td>-88.342</td>
<td>603.889</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>-53.256</td>
<td>172.648</td>
<td>788.89</td>
<td>-392.160</td>
<td>285.648</td>
</tr>
</tbody>
</table>

Table 25: Predicted marginal means from maximal model of centroparietal P600 amplitude elicited by subject- and object-relative clauses.

Figure 30: Scalp array of scaled amplitude of the N420 smoothed with a (100ms) filter.
6.3 *N*420

Object-relatives, compared to subject-relatives, elicited an early, negative ERP component maximal over the left parietal (P3) electrode site consistent with an N400-like effect. To distinguish it from an earlier negative-going deflection with a similar spatial distribution, we will refer to this ERP component as an N420. A scalp array of mean scaled amplitude across group by condition by electrode site for the N420 amplitude is plotted in Figure 30.

Object-relatives elicited an early wave starting around 200ms and peaking around 350ms. Grand mean waveforms of scaled amplitude across group across electrode sites for the N420 amplitude is plotted by condition in Figure 31. Mean scaled amplitude by- and across- group and by condition is reported in Table 26 and plotted in Figure 32.

![Figure 31: Grand mean waveform of N420 amplitude by conditions smoothed with a (100ms) filter.](image)

![Figure 32: Mean N420 amplitude by bilingual group and condition with 95% confidence interval error bars.](image)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Late Bilinguals</th>
<th>Heritage Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-Relatives</td>
<td>-102.58 (493.58)</td>
<td>52.73 (566.84)</td>
<td>-35.71 (531.34)</td>
</tr>
<tr>
<td>Object-Relatives</td>
<td>32.94 (523.26)</td>
<td>-37.72 (553.49)</td>
<td>1.86 (537.30)</td>
</tr>
<tr>
<td>Total</td>
<td>-34.68 (512.66)</td>
<td>6.56 (561.15)</td>
<td>-16.73 (534.38)</td>
</tr>
</tbody>
</table>

*Table 26: Mean N420 amplitude by bilingual group and condition. (Standard deviations in parentheses).*
Contrary to expectations for an N400-like effect, across group, mean scaled amplitude for object-relatives ($M=1.86$, $SD=537.30$) was more positive than for subject-relatives ($M=-35.71$, $SD=531.34$) indicating that object-relatives elicited a weaker effect for the negative-going deflection over posterior electrode sites compared to subject-relatives. Across condition, heritage speakers ($M=6.56$, $SD=561.15$) had more positive mean scaled amplitude than late bilinguals ($M=-34.68$, $SD=512.66$). In looking at mean scaled amplitude, it appears that only heritage speakers exhibited the subject-object processing asymmetry as indexed by a N375. Heritage speakers had more negative mean scaled amplitude for object-relatives ($M=-37.72$, $SD=553.49$) than subject-relatives ($M=52.73$, $SD=566.84$). However, late bilinguals exhibited the opposite effect. They had greater negative scaled amplitude for subject-relatives ($M=-102.58$, $SD=493.58$) than object-relatives ($M=32.94$, $SD=523.26$).

To determine whether differences between condition and bilingual group were statistically significant, N420 amplitude was modeled following the procedure outlined in section 5.3 and visualized in Figure 33. The maximal model, which was also the most parsimonious model, shown in Table 27, retained the predictor variables of group, condition, and the interaction had a significantly better fit than the null model ($\chi^2(3)=12.021$, $p=.007$) and a simpler model without the interaction ($\chi^2(1)=9.681$, $p=.002$).

![Linear mixed-effects models for N420 amplitude. Significant parsimonious model in red type.](image-url)
The intercept of the model was significant, meaning that late bilingual amplitude for subject-relatives significantly differed from the mean across condition and across group, $B=-120.58$, $SE(B)=33.82$, $t(882.80)=-3.033$, $p=.002$. The predictor variable of group was significant, meaning that heritage speakers had significantly higher amplitude for subject-relatives than late bilinguals did for subject-relatives, $B=155.31$, $SE(B)=51.54$, $t(882.80)=3.013$, $p=.003$. The predictor variable of condition was significant, meaning that late bilinguals had significantly higher amplitude for object-relatives than subject-relatives, $B=135.52$, $SE(B)=47.78$, $t(882.80)=2.836$, $p=.005$. Lastly, the interaction between bilingual group and condition was significant, meaning that the difference between subject-relatives and object-relatives was significantly different between bilingual group, $B=-225.97$, $SE(B)=72.43$, $t(882.80)=-3.120$, $p=.002$.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-102.58</td>
<td>33.82</td>
<td>882.80</td>
<td>-3.033</td>
</tr>
<tr>
<td>Group(HS)</td>
<td>155.31</td>
<td>51.54</td>
<td>882.80</td>
<td>3.013</td>
</tr>
<tr>
<td>RC(object)</td>
<td>135.52</td>
<td>47.78</td>
<td>882.80</td>
<td>2.836</td>
</tr>
<tr>
<td>Group:RC</td>
<td>-225.97</td>
<td>72.43</td>
<td>882.80</td>
<td>-3.120</td>
</tr>
</tbody>
</table>

Formula in R: Scaled Amplitude ~ cond * gen + (cond | part) + (1 | item)

† $p<.1$ ‡ $p<.05$ ‡‡ $p<.01$ ‡‡‡ $p<.001$

Table 27: Maximal linear mixed-effects model of N420 amplitude elicited by subject- and object-relative clauses.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>Cl_lower</th>
<th>Cl_upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Late Bilingual</td>
<td>7.509</td>
<td>27.216</td>
<td>882.77</td>
</tr>
<tr>
<td></td>
<td>Heritage Speaker</td>
<td>-24.921</td>
<td>25.769</td>
<td>882.77</td>
</tr>
<tr>
<td>Relative Clause</td>
<td>Subject</td>
<td>-2.388</td>
<td>25.442</td>
<td>882.77</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>-24.921</td>
<td>25.769</td>
<td>882.77</td>
</tr>
<tr>
<td>Late Bilingual</td>
<td>Subject-relatives</td>
<td>-102.575</td>
<td>33.818</td>
<td>882.77</td>
</tr>
<tr>
<td></td>
<td>Object-relatives</td>
<td>32.941</td>
<td>33.749</td>
<td>882.77</td>
</tr>
<tr>
<td>Heritage Speaker</td>
<td>Subject-relatives</td>
<td>52.734</td>
<td>38.922</td>
<td>882.77</td>
</tr>
<tr>
<td></td>
<td>Object-relatives</td>
<td>-37.717</td>
<td>38.082</td>
<td>882.77</td>
</tr>
</tbody>
</table>

Table 28: Predicted marginal means from maximal model of N420 amplitude elicited by subject- and object-relative clauses.

The predicted values from the maximal model were used to calculate descriptive statistics, shown in Table 28 and as outlined in section 5.3, to compare groups and explore the subject-object asymmetry by-and across- groups after having controlled for individual and item level variance. Across condition,
although heritage speakers ($M_p=7.509$, $SE_p=27.216$, 95% CI $[-45.906, 60.924]$) did have higher mean amplitude than late bilinguals ($M_p=-34.817$, $SE_p=223.889$, 95% CI $[-81.702, 12.068]$), this difference was not significant. Similarly, across group and contrary to expectations of object-relatives eliciting greater N420 amplitude, although object-relatives ($M_p=-2.388$, $SE_p=25.442$, 95% CI $[-52.322, 47.546]$) elicited a less negative amplitude than subject-relatives ($M_p=-24.921$, $SE_p=25.769$, 95% CI $[-75.497, 25.656]$), this difference was not significant. While a significant effect of condition was observed in both bilingual groups, heritage speakers were the only group to show the subject-object asymmetry in the expected direction as indexed by the N420. For heritage speakers, object-relatives ($M_p=-37.717$, $SE_p=38.082$, 95% CI $[-112.458, 37.024]$) elicited significantly more negative amplitude than subject-relatives ($M_p=52.734$, $SE_p=38.892$, 95% CI $[-23.597, 129.066]$). For late bilinguals, object-relatives elicited the opposite effect; object-relatives ($M_p=32.941$, $SE_p=33.749$, 95% CI $[-33.297, 99.180]$) elicited significantly more positive amplitude than object-relatives ($M_p=-102.575$, $SE_p=33.818$, 95% CI $[-168.948, -36.202]$).

**Figure 34:** Scalp array of scaled amplitude of the N375 smoothed with a (100ms) filter.
6.3 N375

Object-relatives, compared to subject-relatives, elicited an early, negative ERP component maximal over the left parietal (P3) electrode site consistent with an N400-like effect. To distinguish it from a later negative-going deflection with a similar spatial distribution, we will refer to this ERP component as an N375. A scalp array of mean scaled amplitude across group by condition by electrode site for the N375 amplitude is plotted in Figure 34.

Object-relatives elicited an early wave starting around 325ms and peaking around 375ms. Grand mean waveforms of scaled amplitude across group across electrode sites for the N375 amplitude is plotted by condition in Figure 35. Mean scaled amplitude by- and across- group and condition is reported in Table 29 and plotted in Figure 36.

Figure 35: Grand mean waveform of N375 amplitude by condition smoothed with a (100ms) filter.

Figure 36: Mean N375 amplitude by bilingual group and condition with 95% confidence interval error bars.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Late Bilinguals</th>
<th>Heritage Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-Relatives</td>
<td>-91.45</td>
<td>70.56</td>
<td>-21.69</td>
</tr>
<tr>
<td></td>
<td>(485.26)</td>
<td>(550.37)</td>
<td>(519.93)</td>
</tr>
<tr>
<td>Object-Relatives</td>
<td>35.82</td>
<td>-15.70</td>
<td>13.16</td>
</tr>
<tr>
<td></td>
<td>(500.55)</td>
<td>(531.27)</td>
<td>(514.34)</td>
</tr>
<tr>
<td>Total</td>
<td>-27.68</td>
<td>26.52</td>
<td>-4.09</td>
</tr>
<tr>
<td></td>
<td>(496.58)</td>
<td>(541.71)</td>
<td>(517.11)</td>
</tr>
</tbody>
</table>

Table 29: Mean N375 amplitude by bilingual group and condition. (Standard deviations in parentheses).
Contrary to expectations for an N400-like effect, across group, mean scaled amplitude for object-relatives ($M=13.16, SD=614.34$) was more positive than subject-relatives ($M=-21.69, SD=519.93$) indicating that object-relatives elicited a weaker effect for the negative-going deflection over posterior electrode sites compared to subject-relatives. Across condition, heritage speakers ($M=26.52, SD=541.71$) had more positive mean scaled amplitude than late bilinguals ($M=-27.68, SD=496.58$). In looking at mean scaled amplitude, it appears that only heritage speakers exhibited the subject-object processing asymmetry as indexed by a N375. Heritage speakers had more negative mean scaled amplitude for object-relatives ($M=-15.70, SD=531.27$) than subject-relatives ($M=70.56, SD=550.37$). However, late bilinguals exhibited the opposite effect. They had greater negative scaled amplitude for subject-relatives ($M=-91.45, SD=485.26$) than object-relatives ($M=35.82, SD=500.55$).

To determine whether differences between condition and bilingual group were statistically significant, N375 amplitude was modeled following the procedure outlined in section 5.3 and visualized in Figure 37. The maximal model, which was also the most parsimonious model, shown in Table 30, retained the predictor variables of group, condition, and the interaction and had a significantly better fit than the null model ($\chi^2(3)=12.570, p=.006$) and a simpler model without the interaction ($\chi^2(1)=9.240, p=.002$).

**Figure 37:** Linear mixed-effects models for N375 amplitude. Significant parsimonious model in red type.
## Table 30: Maximal linear mixed-effects model of N375 amplitude elicited by subject- and object-relative clauses.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-91.45</td>
<td>32.71</td>
<td>862.00</td>
<td>-2.795</td>
<td>.005 **</td>
</tr>
<tr>
<td>Group(HS)</td>
<td>162.01</td>
<td>49.86</td>
<td>862.00</td>
<td>3.249</td>
<td>.001 **</td>
</tr>
<tr>
<td>RC(object)</td>
<td>127.27</td>
<td>46.22</td>
<td>862.00</td>
<td>2.754</td>
<td>.006 **</td>
</tr>
<tr>
<td>Group:RC</td>
<td>-213.53</td>
<td>70.06</td>
<td>862.00</td>
<td>-3.048</td>
<td>.002 **</td>
</tr>
</tbody>
</table>

**Formula in R:** Scaled Amplitude ~ cond * gen + (cond | part) + (1 | item)

† p<.1  * p<.05  ** p<.01  *** p<.001

The intercept of the model was significant, meaning that late bilingual amplitude for subject-relatives significantly differed from the mean across condition and across group, $B=-91.45$, $SE(B)=32.71$, $t(862.00)=-2.795$, $p=.005$. The predictor variable of group was significant, meaning that heritage speakers had significantly higher amplitude for subject-relatives than late bilinguals did for subject-relatives, $B=162.01$, $SE(B)=49.86$, $t(862.00)=3.249$, $p=.001$. The predictor variable of condition was significant, meaning that late bilinguals had significantly higher amplitude for object-relatives than subject-relatives, $B=127.27$, $SE(B)=46.22$, $t(862.00)=2.754$, $p=.006$. Critically, the interaction between bilingual group and condition was significant, meaning that the difference between subject-relatives and object-relatives was significantly different between bilingual group, $B=-213.53$, $SE(B)=70.06$, $t(862.00)=-3.048$, $p=.002$.

## Table 31: Predicted marginal means from maximal model of N375 amplitude elicited by subject- and object-relative clauses.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>CI_lower</th>
<th>CI_upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late Bilingual</td>
<td>-27.811</td>
<td>23.109</td>
<td>862.01</td>
<td>-73.168</td>
<td>17.545</td>
</tr>
<tr>
<td>Heritage Speaker</td>
<td>27.431</td>
<td>26.327</td>
<td>862.01</td>
<td>-24.243</td>
<td>79.104</td>
</tr>
<tr>
<td>Relative Clause</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>-10.444</td>
<td>24.928</td>
<td>862.01</td>
<td>-59.371</td>
<td>38.483</td>
</tr>
<tr>
<td>Object</td>
<td>10.064</td>
<td>24.612</td>
<td>862.01</td>
<td>-38.243</td>
<td>58.370</td>
</tr>
<tr>
<td>Late Bilingual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject-relatives</td>
<td>-91.447</td>
<td>32.714</td>
<td>862.01</td>
<td>-155.656</td>
<td>-27.239</td>
</tr>
<tr>
<td>Object-relatives</td>
<td>35.825</td>
<td>32.648</td>
<td>862.01</td>
<td>-28.254</td>
<td>99.904</td>
</tr>
<tr>
<td>Heritage Speaker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject-relatives</td>
<td>70.559</td>
<td>37.623</td>
<td>862.01</td>
<td>-3.283</td>
<td>144.402</td>
</tr>
<tr>
<td>Object-relatives</td>
<td>-15.697</td>
<td>36.839</td>
<td>862.01</td>
<td>-88.001</td>
<td>56.607</td>
</tr>
</tbody>
</table>

The predicted values from the maximal model were used to calculate descriptive statistics, shown in Table 31 and as outlined in section 5.3, to compare groups and explore the subject-object asymmetry by-and across-groups after having controlled for individual and item level variance. Across condition,
heritage speakers \( (M_p=27.431, SE_p=26.327, 95\% \, CI_p [-24.243, 79.104]) \) had significantly higher mean amplitude than late bilinguals \( (M_p=-27.811, SE_p=23.109, 95\% \, CI_p [-73.168, 17.545]) \). Across group and contrary to expectations of object-relatives eliciting greater N420 amplitude, object-relatives \( (M_p=10.064, SE_p=24.612, 95\% \, CI_p [-38.243, 58.370]) \) elicited a more positive amplitude than subject-relatives \( (M_p=-10.444, SE_p=24.928, 95\% \, CI_p [-59.371, 38.483]) \), but this difference was not significant. While a significant effect of condition was observed in both bilingual groups, heritage speakers were the only group to show the subject-object asymmetry in the expected direction as indexed by the N375. For heritage speakers, object-relatives \( (M_p=-15.697, SE_p=36.839, 95\% \, CI_p [-88.001, 56.607]) \) elicited significantly more negative amplitude than subject-relatives \( (M_p=70.559, SE_p=37.623, 95\% \, CI_p [-3.283, 144.402]) \). For late bilinguals, object-relatives elicited the opposite effect; object-relatives \( (M_p=35.825, SE_p=32.648, 95\% \, CI_p [-28.254, 99.904]) \) elicited significantly more positive amplitude than subject-relatives \( (M_p=-91.447, SE_p=32.714, 95\% \, CI_p [-155.656, -27.239]) \).

### Table 32: ERP component correlations and regression coefficients with 95% confidence intervals across bilingual group

<table>
<thead>
<tr>
<th>P600 effects</th>
<th>Correlation</th>
<th>95% CI</th>
<th>Coefficient</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal P600 ~ CP P600</td>
<td>-.20</td>
<td>[-.51, .15]</td>
<td>-.40</td>
<td>[-.46, -.34]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biphasic effects</th>
<th>Correlation</th>
<th>95% CI</th>
<th>Coefficient</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal P600 ~ N375</td>
<td>-.02</td>
<td>[-.36, .32]</td>
<td>-.18</td>
<td>[-.25, -.12]</td>
</tr>
<tr>
<td>Frontal P600 ~ N420</td>
<td>.00</td>
<td>[-.34, .34]</td>
<td>-.18</td>
<td>[.24, .11]</td>
</tr>
<tr>
<td>CP P600 ~ N375</td>
<td>.93***</td>
<td>[.86, .96]</td>
<td>.89</td>
<td>[.85, .91]</td>
</tr>
<tr>
<td>CP P600 ~ N420</td>
<td>.94***</td>
<td>[.88, .97]</td>
<td>.88</td>
<td>[.85, .91]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N400 effects</th>
<th>Correlation</th>
<th>95% CI</th>
<th>Coefficient</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>N420~N375</td>
<td>.98***</td>
<td>[.96, .99]</td>
<td>.96</td>
<td>[.94, .98]</td>
</tr>
</tbody>
</table>

6.5 Across ERP components

Across bilingual group, the correlations between ERP component difference amplitude can be seen in Table 32. The frontal P600 was not significantly correlated with any of the other ERP component. There was a non-significant, weak, negative correlation between the frontal P600 and the centroparietal P600, \( r(32)=-.20, 95\% \, CI [-.51, .15], p=.25 \). The frontal P600 was neither significantly correlated with the N375 \( (r(32)=-.02, 95\% \, CI [-.36, .32], p=.90) \) nor the N420 \( (r(32)=.003, 95\% \, CI [-.34, .34], p=.99) \). There was a significant, strong, positive correlation between the centroparietal P600 and both the N375 \( (r(32)=.93, 95\% \, CI [.86, .96], p<.001) \) and the N420 \( (r(32)=.94, 95\% \, CI [.88, .97], p<.001) \), respectively. There was a
significant, strong, positive correlation between the two N400-like effects, $r(32)=.98$, 95% CI [0.96, 0.99], $p<.001$. The N400-like effects were significantly correlated with each other more strongly than either N400-like effect was with the centroparietal P600.

Across bilingual group, the regression coefficients for ERP component difference amplitude predicting other ERP component difference amplitude can be seen in Table 32. The centroparietal P600 significantly predicted the frontal P600, $\chi^2(1)=153.300$, $p<.001$. An increase in centroparietal P600 difference amplitude is associated with a significant decrease in frontal P600 difference amplitude, $B=-0.401$, SE($B$)=0.031, $t(870.00)=-12.691$, $p<.001$. The N375 significantly predicted the frontal P600, $\chi^2(1)=29.694$, $p<.001$. An increase in N375 difference amplitude is associated with a significant decrease in frontal P600 difference amplitude, $B=-0.183$, SE($B$)=0.033, $t(871.70)=-5.502$, $p<.001$. The N420 significantly predicted the frontal P600, $\chi^2(1)=27.326$, $p<.001$. An increase in N420 difference amplitude is associated with a significant decrease in frontal P600 difference amplitude, $B=-0.176$, SE($B$)=0.033, $t(871.00)=-5.276$, $p<.001$. The centroparietal P600 difference amplitude was associated with a significantly greater decrease in frontal P600 difference amplitude than either of the N400-like effects.

The N375 significantly predicted the centroparietal P600, $\chi^2(1)=1287.300$, $p<.001$. An increase in N375 difference amplitude is associated with a significant increase in centroparietal P600 difference amplitude, $B=0.878$, SE($B$)=0.016, $t(864.90)=54.267$, $p<.001$. The N420 significantly predicted centroparietal P600, $\chi^2(1)=1319.700$, $p<.001$. An increase in N420 difference amplitude is associated with a significant increase in centroparietal P600 difference amplitude, $B=0.883$, SE($B$)=0.016, $t(861.40)=55.708$, $p<.001$.

There was not a significant difference between N400-like effects in predicting the centroparietal P600. The N375 significantly predicted the N420, $\chi^2(1)=2260.700$, $p<.001$. An increase in N375 difference amplitude is associated with a significant increase in N420 difference amplitude, $B=0.962$, SE($B$)=0.009, $t(873.00)=103.700$, $p<.001$. The N400-like effects were more significant predictors of each other than either of the N400-like effects predicting the centroparietal P600.

For late bilinguals, the correlations between ERP component difference amplitude can be seen in Table 33. The frontal P600 was not significantly correlated with any of the other ERP component. There was a non-significant, weak, negative correlation between the frontal P600 and the centroparietal P600, $r(17)=-.13$, 95% CI [-.55, .35], $p=.60$. There was a non-significant, weak, positive correlation between the
frontal P600 and both the N375 ($r(17)$=.19, 95% CI [-.28, .60], $p$=.42), and the N420 ($r(17)$=.23, 95% CI [-.25, .62], $p$=.34), respectively. There was a significant, strong, positive correlation between the centroparietal P600 and both the N375 ($r(17)$=.80, 95% CI [.55, .92], $p$<.001) and the N420 ($r(17)$=.83, 95% CI [.60, .93], $p$<.001), respectively. There was a significant, strong, positive correlation between the two N400-like effects, $r(17)$=.97, 95% CI [.91, .99], $p$<.001. The N400-like effects were significantly correlated with each other more strongly than either N400-like effect was with the centroparietal P600.

<table>
<thead>
<tr>
<th>P600 effects</th>
<th>Correlation</th>
<th>95% CI</th>
<th>Coefficient</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal P600 ~ CP P600</td>
<td>-.13</td>
<td>[-.55, .35]</td>
<td>-.36</td>
<td>[-.44, -.28]</td>
</tr>
<tr>
<td>Biphasic effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal P600 ~ N375</td>
<td>.19</td>
<td>[-.28, .60]</td>
<td>-.13</td>
<td>[-.22, -.04]</td>
</tr>
<tr>
<td>Frontal P600 ~ N420</td>
<td>.23</td>
<td>[-.25, .62]</td>
<td>-.12</td>
<td>[-.21, -.04]</td>
</tr>
<tr>
<td>CP P600 ~ N375</td>
<td>.80***</td>
<td>[.55, .92]</td>
<td>.88</td>
<td>[.84, .92]</td>
</tr>
<tr>
<td>CP P600 ~ N420</td>
<td>.83***</td>
<td>[.60, .93]</td>
<td>.89</td>
<td>[.85, .93]</td>
</tr>
<tr>
<td>N400 effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N420~N375</td>
<td>.97***</td>
<td>[.91, .99]</td>
<td>.97</td>
<td>[.95, .99]</td>
</tr>
</tbody>
</table>

**Table 33:** ERP component correlations and regression coefficients with 95% confidence intervals for late bilinguals

For late bilinguals, the regression coefficients for ERP component difference amplitude predicting other ERP component difference amplitude can be seen in Table 33. The centroparietal P600 significantly predicted the frontal P600, $\chi^2(1)$=69.101, $p$<.001. An increase in centroparietal P600 difference amplitude is associated with a significant decrease in frontal P600 difference amplitude, $B$=-0.362, $SE(B)=0.042$, $t(493.00)=-8.676$, $p$<.001. The N375 significantly predicted the frontal P600, $\chi^2(1)$=8.658, $p$<.01. An increase in N375 difference amplitude is associated with a significant decrease in frontal P600 difference amplitude, $B$=-0.132, $SE(B)=0.044$, $t(493.00)=-2.965$, $p$<.01. The N420 significantly predicted the frontal P600, $\chi^2(1)$=7.676, $p$<.01. An increase in N420 difference amplitude is associated with a significant decrease in frontal P600 difference amplitude, $B$=-0.124, $SE(B)=0.044$, $t(492.70)=-2.788$, $p$<.01. Centroparietal P600 difference amplitude was associated with a significantly greater decrease in frontal P600 difference amplitude than either of the N400-like effects. The N375 significantly predicted the centroparietal P600, $\chi^2(1)$=728.100, $p$<.001. An increase in N375 difference amplitude is associated with a significant increase in centroparietal P600 difference amplitude, $B$=0.881, $SE(B)=0.022$, $t(492.40)=40.817$, $p$<.001. The N420 significantly predicted the centroparietal P600, $\chi^2(1)$=758.350,
An increase in N420 difference amplitude is associated with a significant increase in centroparietal P600 difference amplitude, $B=0.887$, $SE(B)=0.021$, $t(489.90)=42.482$, $p<.001$. There was not a significant difference between N400-like effects in predicting the centroparietal P600. The N375 significantly predicted the N420, $\chi^2(1)=1364.400$, $p<.001$. An increase in N375 difference amplitude is associated with a significant increase in N420 difference amplitude, $B=0.969$, $SE(B)=0.011$, $t(493.00)=85.763$, $p<.001$. The N400-like effects were more significant predictors of each other than either of the N400-like effects were in predicting the centroparietal P600.

<table>
<thead>
<tr>
<th>P600 effects</th>
<th>Correlation</th>
<th>95% CI</th>
<th>Coefficient</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal P600 ~ CP P600</td>
<td>-.44†</td>
<td>[-.78, .09]</td>
<td>-.45</td>
<td>[-.54, -.36]</td>
</tr>
<tr>
<td>Biphasic effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal P600 ~ N375</td>
<td>-.34</td>
<td>[-.72, .21]</td>
<td>-.24</td>
<td>[-.34, -.14]</td>
</tr>
<tr>
<td>Frontal P600 ~ N420</td>
<td>-.32</td>
<td>[-.72, .23]</td>
<td>-.24</td>
<td>[-.33, -.14]</td>
</tr>
<tr>
<td>CP P600 ~ N375</td>
<td>.96***</td>
<td>[.89, .99]</td>
<td>.88</td>
<td>[.83, .92]</td>
</tr>
<tr>
<td>CP P600 ~ N420</td>
<td>.97***</td>
<td>[.91, .99]</td>
<td>.88</td>
<td>[.83, .93]</td>
</tr>
<tr>
<td>N400 effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N420 ~ N375</td>
<td>.98***</td>
<td>[.95, .99]</td>
<td>.95</td>
<td>[.92, .98]</td>
</tr>
</tbody>
</table>

Table 34: ERP component correlations and regression coefficients with 95% confidence intervals for heritage speakers

For heritage speakers, the correlations between ERP component difference amplitude can be seen in Table 34. The frontal P600 was not significantly correlated with any of the other ERP component. The medium, negative correlation between the frontal P600 and the centroparietal P600 approached significance, $r(13)=-.44$, 95% CI [-.78, .09], $p=.10$. There was a non-significant, medium, negative correlation between the frontal P600 and both the N375 ($r(13)=-.34$, 95% CI [-.72, .21], $p=.22$) and the N420 ($r(13)=-.32$, 95% CI [-.72, .23], $p=.24$), respectively. There was a significant, strong, positive correlation between the centroparietal P600 and both the N375 ($r(13)=.96$, 95% CI [.89, .99], $p<.001$) and the N420 ($r(13)=.97$, 95% CI [.91, .99], $p<.001$), respectively. There was a significant, strong, positive correlation between the two N400-like effects, $r(13)=.98$, 95% CI [.95, .99], $p<.001$.

For heritage speakers, the regression coefficients for ERP component difference amplitude predicting other ERP component difference amplitude can be seen in Table 34. The centroparietal P600 significantly predicted the frontal P600, $\chi^2(1)=83.626$, $p<.001$. An increase in centroparietal P600 difference amplitude is associated with a significant decrease in frontal P600 difference amplitude, $B=-$.
The N375 significantly predicted the frontal P600, $\chi^2(1)=22.373$, $p<.001$. An increase in N375 difference amplitude is associated with a significant decrease in frontal P600 difference amplitude, $B=-0.240$, $SE(B)=0.050$, $t(378.40)=-4.829$, $p<.001$. The N400 significantly predicted the frontal P600, $\chi^2(1)=21.589$, $p<.001$. An increase in N400 difference amplitude is associated with a significant decrease in frontal P600 difference amplitude, $B=-0.235$, $SE(B)=0.050$, $t(375.10)=-4.730$, $p<.001$. Centroparietal P600 difference amplitude was associated with a significantly greater decrease in frontal P600 difference amplitude than either of the N400-like effects. The N375 significantly predicted the centroparietal P600, $\chi^2(1)=559.55$, $p<.001$. An increase in N375 difference amplitude is associated with a significant increase in centroparietal P600 difference amplitude, $B=0.875$, $SE(B)=0.025$, $t(379.20)=35.735$, $p<.001$. The N420 significantly predicted the centroparietal P600, $\chi^2(1)=565.940$, $p<.001$. An increase in N420 difference amplitude is associated with a significant increase in centroparietal P600 difference amplitude, $B=0.879$, $SE(B)=0.024$, $t(376.70)=36.145$, $p<.001$. There was not a significant difference between N400-like effects in predicting the centroparietal P600. The N375 significantly predicted the N420, $\chi^2(1)=922.710$, $p<.001$. An increase in N375 difference amplitude is associated with a significant increase in N420 difference amplitude, $B=0.953$, $SE(B)=0.015$, $t(380.00)=62.678$, $p<.001$. The N400-like effects were more significant predictors of each other than either of the N400-like effects were in predicting the centroparietal P600.

The correlations between N400-like effects and the centroparietal P600 were significantly greater for heritage speakers than late bilinguals, however, there is no significant difference between N400-like effects and the centroparietal P600 patterns between bilingual groups. Additionally, there was no significant between-group differences for the N400-like effects. Likewise, the patterns seen between the N400-like effects and the frontal P600 were in a similar direction across bilingual group, although the effect was stronger in heritage speakers than late bilinguals. The frontal P600 was predicted significantly better by the centroparietal P600, the N375, and the N420 for heritage speakers than late bilinguals.

7. Discussion
The present study found that heritage speakers process relative clauses differently than their time-apparent parents, late bilinguals. For all linguistic ERP components isolated with the PCA, heritage
speakers patterned differently than late bilinguals. Previous ERP literature identified three ERP components that indexed processing difficulty of relative clauses: LAN, N400, and P600. The ERP component effects by bilingual group in the current study are in Table 35. Significant, observed effects are indicated with a tick mark <✓>. No LAN-like effect was observed. Two possible N400-like ERP components were observed and significant. For both late bilinguals and heritage speaker bilinguals amplitude for object-relatives was significantly different than subject-relatives. However, for both N400-like ERP components, heritage speakers and late bilinguals differed in the direction of the effect. For heritage speakers, object-relatives had significantly more negative amplitude than subject-relatives; this is in the expected direction for N400-like effects. However, for late bilinguals, object-relatives had significantly more positive amplitude than subject-relatives; this is in opposite direction that is expected for N400-like effects. For the two P600-like ERP components, a significant effect was observed only for the centroparietal P600 and not for the frontal P600. A significant centroparietal P600 effect was found only for late bilinguals. Heritage speakers did not process object-relatives differently than subject-relatives as indexed by the centroparietal P600. Additional findings of note that will be further discussed below are that, for the frontal P600, heritage speakers had greater amplitude than late bilinguals, although there was no significant subject-object relative clause processing asymmetry observed, and, for the centroparietal P600, subject-relative amplitude for heritage speakers was significantly greater than subject-relative amplitude for late bilinguals.

<table>
<thead>
<tr>
<th>ERP Component</th>
<th>Late Bilinguals</th>
<th>Heritage Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAN</td>
<td>LAN</td>
<td>--</td>
</tr>
<tr>
<td>N400</td>
<td>N420</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>N375</td>
<td>✓</td>
</tr>
<tr>
<td>P600</td>
<td>Frontal P600</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Centroparietal P600</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 35: ERP component results

ERP component amplitude for the N400-like effects and the centroparietal P600 effect seem to be driven by a similar underlying processes. The frontal P600 was not highly correlated with the centroparietal P600.
or the two N400-like ERP components and likely indexes a different processing function. The two N400-like ERP components were significantly, strongly correlated and were significant predictors of each other. This explains the similar results for the two early negative-going deflections. The centroparietal P600 was strongly correlated with the two N400-like ERP components. Higher N400-like amplitude was associated with significantly higher centroparietal P600 amplitude.

The first research question was: Are ERP components that indicate processing difficulty greater for object-relatives than subject-relatives in late bilinguals? We can confirm our initial hypothesis that object-relative clauses were indeed processed differently than subject-relative clauses in late bilinguals as indexed by the significant N420, N375, and centroparietal P600 effects. However, there was no significant subject-object relative clause processing difference as indexed by frontal 600 or LAN effects.

The second research question was: Do heritage speakers process Spanish object-relatives and subject-relatives differently from late bilinguals? We cannot confirm our initial hypothesis that heritage speakers will process relative clauses similarly to late bilinguals. In all cases, heritage speakers differed from late bilinguals in processing the relative clauses. For the N400-like effects, the significant effect was in the opposite direction for heritage speakers than for late bilinguals. For the centroparietal P600 effect, heritage speakers also differed from late bilinguals in that there was no significant effect. Additionally, heritage speakers had significantly higher P600 amplitude for the subject-relatives than late bilinguals. For the frontal P600, while neither group exhibited a P600 effect, there were still group differences: heritage speakers had significantly higher frontal P600 amplitude than late bilinguals. However, in examining the significant correlations between ERP components, heritage speakers broadly did have similar ERP response structure (related N375, N420, and centroparietal P600 effects), but differed in the relationship of those ERP responses to the frontal P600 component (stronger relationship between the N375, N420, and centroparietal P600 effects to the frontal P600 effects for heritage speakers than late bilinguals). The lack of a P600 effect for heritage speakers and the opposing N400-like effects in heritage speakers and late bilinguals are further discussed. Possible sources of group differences are explored and discussed below.
7.1 Comparison to previous studies

Some researchers (King & Kutas, 1995; Mecklinger et al., 1995; Müller, King, & Kutas, 1997; Ueno & Garnsey, 2007; Yang, Perfetti, & Liu, 2010; Kwon et al., 2013; Wang et al., 2015) found LAN or N400 effects indexed the subject-object relative processing asymmetry, although the spatial distribution of these negativities is broader and less localized than proto-typical LANs or N400s in that they are not restricted, or even maximal, over left anterior or centroparietal electrode sites, respectively. The present study found early, posterior, negative-going deflections, N400 effects, only for heritage speakers consistent with some of the previous literature. All ERP effects observed were left lateralized and none of the negative deflections were anterior. No LAN effect as observed in the present study as in previous studies.

Other researchers (Mecklinger et al., 1995; Ueno & Garnsey, 2007; Yang, Perfetti, & Liu, 2010; Wang et al., 2015) have reported biphasic N400/LAN-P600 effects. Neither of the groups in the present study elicited statistically significant biphasic N400-P600 effects. Lastly, some researchers (Carreiras et al., 2010) observed only P600 effects for more difficult to process relative clauses. A centroparietal P600 was found for late bilinguals in response to harder to process relative clauses, consistent with the previous literature.

The frontal P600, while not previously reported as an index of the subject-object relative clause processing asymmetry, was observed in heritage speakers across sentence type. The frontal P600 is an index of syntactic ambiguity resolution (Hagoort et al., 1999; Kaan & Swaab, 2003; Ledoux et al., 2007) or syntactic integration difficulties with complex sentences (Friederici, Hahne, & Saddy, 2002). The frontal P600 effect, while not previously reported in the relative clause processing literature, is consistent with the idea that heritage speakers are indeed processing complex sentences and might be experiencing more difficulty when processing than late bilinguals.

7.2 Temporo-spatial PCA

The ERP components were isolated for hypothesis testing by a sequential temporo-spatial PCA similar to Spencer, Dien, and Donchin (2001), Dien and Frishkoff (2005) and Hestvik and Durvasula (2016). This analysis uses a data-driven “blind-source separation” approach that aims to identify and isolate the
underlying componential structure of the ERP responses while limiting researcher bias. A benefit of this analysis is that ERP components can be isolated which might not be apparent in visual inspection of voltage data. However, the analysis is very underutilized in sentence processing ERP studies making comparisons to previous literature not very straightforward. Traditional ERP analyses are conducted by examining the voltage data and selecting a number of electrodes to average over in a given time window. The spatial PCA, a PCA with an oblique rotation, specifically models the interrelatedness of neuroelectrical activity in a way that traditional analyses do not. A consequence of using the temporal PCA to establish gradient time windows of interest and to model the periodic oscillation of neuroelectrical activity in addition to the spatial PCA results in ERP components that are similar to, but slightly different from previously discussed ERP components. Nothing is to be inferred from differences in spatial or temporal distribution for the PCA-isolated ERP components in the present study. The differences arise from the different isolation procedures and further study is needed to establish whether the differences are consequential (such as by using data from previously published studies and reanalyzing using a temporo-spatial PCA). The PCA is a more fine-grained sorting of ERP components, as what might appear to be one deflection might be composed of a number of faster deflections leading to the appearance of a slower, sustained deflection. In the case of the present study, two deflections consistent with an N400 were observed, the N375 and the N420, in that there were negative-going deflections peaking between 300ms and 500ms. Given the similar topographic distribution, both deflections would show up as a longer, negative-going deflection. Therefore, we argue that the N375 and N420 are both examples of, but more likely constituent components making up, the traditional N400 reported in previous literature. We are agnostic to which specific and distinct processes they index. Similarly, the PCA was conducted to isolate components broadly consistent with the latency, polarity, and scalp distribution of linguistic ERP components in the literature. However, since the process of isolating the ERP components are different, the ERP components in the present study do not look exactly like those previously reported. For example, the centroparietal P600 observed in the present study showed a slow, late wave that was maximal over centroparietal (C3, Cz, CP3, CPz, Pz) electrode sites, consistent with previous literature, but also showed a slow, late, positive wave over frontoparietal (FP1, FP2) electrode sites. In a traditional approach, the amplitude from only the centroparietal electrode sites would be included in the analysis, but in the current
approach, the spatial PCA identifies that positive amplitude over centroparietal regions co-occurs over frontoparietal regions.

In the present study, the temporo-spatial PCA was conducted over all participants and not by-group or by-individual therefore the ERP components isolated are the same for both late bilinguals and heritage speakers. The PCA in the present study doesn't assume any temporal or spatial distributional differences between the two bilingual groups. The difference between groups that is expected is assumed to be in amplitude only and not in the componential distribution. This assumption might be inappropriate, but given the great variability, both temporal and spatial, observed in bilinguals (e.g., Kotz et al., 2008; Moreno & Kutas, 2005) and even monolinguals (e.g., Weber et al., 2003; Pakulak & Neville, 2010), this across group PCA was a first pass in understanding processing in the first-learnt language of bilinguals living in an L1/ML society. Indeed, Birdsong (2018) argues that great variability is an “inherent characteristic” of bilingualism. Future studies will run a temporo-spatial PCA by group to characterize the early and late wave topographic and latency differences between heritage speakers and late bilinguals, however it is unclear if such an analysis will find differences, regardless of how small, due to the a priori grouping of participants. It would be difficult to understand the important of small differences in such an analysis, such as slightly different weightings on each of the 250 time points or electrodes. The great variability expected suggests that such an analysis would find distributional differences for which there are straightforward ways of testing for significance or conceptualizing in terms of meaningfulness. Conceptually, it is possible to first run the PCA analysis for late bilinguals, then fit the results to the heritage speakers and look for differences in weight as well as goodness of fit of the final PCA solution. However, the temporo-spatial PCA in the present study is a sequential, two-part analysis which makes a confirmatory factor analysis difficult. The goodness of fit of a PCA solution could be tested for the temporal PCA but without the spatial distribution, it would be unclear that the goodness of fit corresponds to in terms of discrete ERP components. This same variability, high temporal and spatial variability between participants, precluded a by-participant temporo-spatial PCA as results were indeterminate. The inter-participant variability resulted in difficulty in the determination of whether individual components corresponded to the experimental manipulations.
7.3 Lack of P600-effect for heritage speakers

For the centroparietal P600, the effect observed for late bilinguals is consistent with previous literature. As anticipated, greater amplitude is found in response to the object-relatives than subject-relatives, reflecting processing difficulty arising from syntactic reanalysis and/or maintaining the filler in working memory. No difference in centroparietal P600 amplitude between object- and subject-relatives was observed for heritage speakers which indicates that they are processing these constructions differently than late bilinguals.

One possible reason for the lack of an effect for heritage speakers is that they could have abandoned processing of the object-relatives. If they had abandoned processing, then no effect of syntactic reanalysis or maintenance of the filler in working memory would be observed, similar to the results from Epstein and colleagues (2013) where no effect was found for children with specific language impairment (SLI). This was argued to be a result of possible disengagement with the task as a larger effect was anticipated given the increased difficulty in processing object wh-questions relative to subject wh-questions. Task disengagement can be ruled out as significant N400 effects were found for heritage speakers and strong correlations between the observed ERP components indicate that heritage speakers were still processing the relative clauses.

A more likely reason for the lack of a P600 effect for heritage speakers is that they have more difficulty processing relative clauses in general. Note that a more positive amplitude for subject-relatives was observed in heritage speakers compared to late bilinguals. If object-relatives elicit a higher amplitude due to syntactic reanalysis or working memory demands, and subject-relatives also elicit a higher amplitude for another reason, such as processing difficulty, then no difference in amplitude would be observed. King and Kutas (1995) divided participants into “good” and “poor” comprehenders and observed greater ERP responses for “poor” comprehenders, although the ERP response was a posterior negativity. They attributed the greater effect for “poor” comprehenders to a related to integration difficulty. “Good” comprehenders, they argue, are better at syntactic and thematic integration; therefore they do not show large integration effects relative to “poor” comprehenders. Extending this argument to the present study, late bilinguals are better at integration of the filler-gap dependency for subject-relatives and therefore have lower amplitude for subject-relatives. If integration is more difficult for heritage speakers,
possibly a result of their less frequent HL use and different patterns of L1 exposure from late bilinguals, then they would have higher P600 amplitude for subject-relatives and not show an asymmetry in amplitude relative to object-relatives.

Further evidence in support of greater processing difficulty associated with higher P600 amplitude comes from ERP studies of bilinguals. Jessen and colleagues (2017) observed greater P600 effects for non-native speakers when processing English fronted indirect objects in a filler-gap paradigm and Rossi and colleagues (2006) observed greater P600 effects for highly proficient L2 speakers for word category violations. Jessen and colleagues (2017) argued that this reflects increased brain activity in L2 processing due to greater processing costs in line with fMRI studies (Wartenburger et al., 2003; Rüschemeyer et al., 2006). In addition to higher subject-relative centroparietal P600 amplitude, heritage speakers also had higher frontal P600 amplitude across condition, possibly reflecting greater processing difficulty of relative clauses generally. Kaan and colleagues (2000) argued that the P600 component does not only reflect syntactic processing, but integration difficulty in general, even in grammatical sentences. In examining the P600 effects, heritage speakers have more difficulty in processing relative clauses generally which explains the lack of a P600 effect indexing only a subject-object relative clause processing asymmetry.

An alternative explanation for the lack of a centroparietal P600 effect comes from taking into account the significant relationship between the N400-like effects and the centroparietal P600 effect. There was a strong positive correlation between these components. More negative N400-like effects were observed for heritage speakers which attenuate later positive P600 amplitude (Brouwer et al., 2016; Brouwer & Crocker, 2017). A stronger significant N400-like effect resulting from surprisal at encountering a noun phrase in subject position for object-relatives would attenuate P600 effects resulting from structural reanalysis or increased working memory demands.

7.4 Opposing N400-like effects

Conflicting results for the N400-like effects were found for heritage speakers and late bilinguals. For heritage speakers, N400-like effects were observed that were in the expected direction: greater negative amplitude for object-relatives compared to subject-relatives. For late bilinguals, the N400-like effects were significant as well, but in the opposite direction than was expected: greater positive amplitude for object-
relatives compared to subject-relatives. The late bilingual positivity is surprising as early positive-going deflections were not found in previous ERP research on relative clause processing. A few conclusions are possible: 1) late bilinguals and heritage speakers are processing relative clauses using different strategies as indexed by distinct ERP components; 2) the effect is indeed an N400 and this effect is non-significant in late bilinguals; or 3) the effect is an early positivity, a P375 and P420, and this effect is non-significant in heritage speakers.

The significant correlations and mixed-effects models exploring the relationship between the two N400-like effects and the centroparietal effects suggest that the third conclusion is correct. The N400-like effects are instead positive-going deflections peaking at 375ms and 420ms, respectively, and this effect was significant in late bilinguals only. Regardless of the significant direction of the effect, the N400-like effects were significantly, positively correlated with the centroparietal P600 by- and across- bilingual group. This effect was significantly stronger for heritage speakers than it was for late bilinguals. There was no significant difference between bilingual group in the N400-like effects predicting centroparietal P600 amplitude. Likewise, the very strong, significant correlations and regression coefficients between both P375 and P420 effects indicate that they are very similar and likely are manifestations of the same process.

Further evidence in support of the early, positive P375 and P420 comes from literature observing P300 or P400 ERP components in response to unexpected stimuli. Early, positive-going deflections have been observed in response to violations of sequential expectancy as a P400 (Dien et al., 2010), rare events detected while sleeping as a P400 (Pentikäinen, 2016), detection of a rule violation as a P300 (Monte-Ordoño & Toro, 2017), and reorganization/revision of one’s mental model as a P300 (Duncan-Johnson & Donchin, 1977; Donchin, 1981; Aleksandrov & Maksimova, 1985; Donchin & Coles, 1988). Indeed, the P600 was initially argued to belong to the P300 family of positive components observed in response to unexpected stimuli (Osterhout & Holcomb, 1992). In the case of the present study, it appears that P400-like effects were observed in response to the unexpected object-gap, whereas no P400-like effects were observed in heritage speakers. The lack of a P400-like effect could be explained by the greater processing difficulty for heritage speakers. The processing difficulty is possibly related to higher working memory demands which have been shown to attenuate P300 responses (Polich, 2007; Evans &
Pollack, 2011), or to the effect of task difficulty which have also been shown to reduce P300 amplitude for harder tasks (Danker et al., 2008).

An open question remains then as to why N400 effects were observed in heritage speakers in the absence of P300/P400 responses, which would have masked any N400 effects. One possible explanation is that N400 effects are also present in late bilinguals and are masked by the P600 and earlier P300/P400 effects. In heritage speakers, the N400 is not being masked by the synchronous and later sustained positivities. This explanation places the locus of processing differences on the cognitive processes indexed by the positive-going deflections. In heritage speakers, object relative clauses might elicit an N400 effect for surprisal at encountering a filled gap or for working memory demands but fail to undergo syntactic reanalysis until later in processing. Late bilinguals also might have an N400 for surprisal at encountering a filled gap or for working memory demands but the N400 might be masked if late bilinguals then immediately start to syntactically reanalyze the structure as indexed by the P300/P400 and P600.

7.5 Heritage speaker processing difficulty
Although heritage speakers are native L1 speakers of Spanish, greater processing difficulty observed for heritage speakers as indexed by the attenuated P400, and greater frontal and centroparietal P600 effects could be attributed to differences in language use and exposure. Indeed, heritage speakers scored at ceiling on the Spanish RMST (Klein & Martohardjono, 2009), which included relative clause structures indicating competence in Spanish syntax. The heritage speakers in the present study, and in previous studies, vary from late bilinguals in terms of language ability (self-report as not proficient users), language use (use Spanish in different domains), and language exposure (Spanish is acquired under different contexts). These differences could be the underlying reason for heritage speakers increased processing difficulty. Martohardjono and colleagues (2017a, 2017b) and Tanner and colleagues (2009) found that language exposure and language use modulated ERP component amplitude in response to ungrammaticality. Differences between English language exposure, current Spanish language use, and self-reported language abilities are explored further.
Three language exposure variables looked at self-reported exposure to English cumulatively over the participants' lifetime, English-medium education, and English used in the community of the participant from ages 5-18. Heritage speakers reported significantly more exposure to English than late bilinguals, as can be seen in Table 36. Cumulative English exposure is operationalized as a percentage of participant’s life spent in an English-dominant environment (either starting a school where English is the language of learning and teaching or age of arrival to the mainland anglophone US). In terms of a percentage of the participant’s life, heritage speakers (M=86.67%, SD=11.26) reported being in an English-dominant environment for a significantly larger percentage of their life than late bilinguals (M=20.00%, SD=17.88), t(31)=-12.29, p<.001, r=.91. A significantly larger percentage of heritage speakers’ education (M=92.35%, SD=10.47) was in English-medium classrooms than late bilinguals (M=13.72%, SD=31.32), t(23.14)=−10.20, p<.001, r=.90. Similarly, a significantly larger percentage of heritage speakers’ youth (M=30.87%, SD=36.14) between the ages of 5 and 18 were spent in English-dominant local communities (or neighborhoods) than late bilinguals (M=1.88%, SD=4.67), t(13.32)=−2.98, p<.05, r=.63.

Table 36: Mean percent of English use for language exposure variables (Standard deviations in parentheses; Significance of group differences from t-test indicated)
There were also group differences in the percentage of Spanish participants reported currently using. Although, in certain domains, both groups used equivalent amounts of Spanish, as can be seen in Table 37. Heritage speakers ($M=64.04\%$, $SD=13.75$) reported using significantly less Spanish with family members (i.e., mother, father, siblings, younger children, partner) than late bilinguals ($M=87.72\%$, $SD=9.45$), $t(31)=5.87$, $p<.001$, $r=.73$. This trend also extends to non-family members (i.e., friends, boss, co-workers, classmates). Heritage speakers ($M=21.01\%$, $SD=13.87$) reported using significantly less Spanish with non-family members than late bilinguals ($M=40.20\%$, $SD=26.29$), $t(28.53)=2.70$, $p<.05$, $r=.45$. Both groups reported using Spanish in the home, in social settings and at work as well as when listening to music, reading, and watching TV at equivalent levels. Percentage of Spanish use in the home, in social settings, and at work did not differ significantly between heritage speakers ($M=49.21\%$, $SD=17.10$) and late bilinguals ($M=41.50\%$, $SD=26.85$), $t(31)=-0.94$, $p=.35$, $r=.17$. Likewise, the percentage of the time that anglophone US Latinx bilinguals read in Spanish, watched Spanish-language TV, and listened to Spanish-language music did not significantly differ (Heritage speakers: $M=72.00\%$, $SD=19.63$; Late bilinguals: $M=62.27\%$, $SD=28.06$), $t(31)=-1.11$, $p=.28$, $r=.20$.

The groups also differed in self-reported Spanish and English language ability in Table 38, although this is not surprising given the well-documented effects of linguistic insecurity in heritage speakers. Participants were asked to rate their language ability on a 5-point Likert scale with the endpoints labeled 1=limited knowledge and 5=native. Late bilinguals reported higher Spanish proficiency in both literacy and speaking. Late bilinguals ($M=4.95$, $SD=0.23$) self-reported significantly higher Spanish literacy skills than heritage speakers ($M=3.97$, $SD=0.84$), $t(14.42)=4.23$, $p<.001$, $r=.74$. All late bilinguals reported their Spanish speaking skills as native. Heritage speakers, however, reported the Spanish speaking skills ($M=4.68$, $SD=0.54$) as significantly different than native, $t(13)=-2.22$, $p<.05$, $r=.53$. The opposite was true for English skills. Heritage speakers (Literacy: $M=4.81$, $SD=0.38$; Speaking: $M=4.71$, $SD=0.47$) reported significantly more native-like English skills than late bilinguals (Literacy: $M=4.08$, $SD=0.97$; Speaking: $M=3.89$, $SD=0.59$), $t(29.38)=-3.98$, $p<.001$, $r=.59$; Speaking: $t(31)=-4.28$, $p<.001$, $r=.60$. 
<table>
<thead>
<tr>
<th>Language Ability</th>
<th>Group</th>
<th>Late Bilinguals (%)</th>
<th>Heritage Speakers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(Standard deviations in parentheses; Significance of group differences from t-test indicated; 1=limited knowledge, 5=native)</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>Literacy</td>
<td>4.95 (0.23)</td>
<td>3.97 (0.84)</td>
</tr>
<tr>
<td></td>
<td>Speaking</td>
<td>5.00 (0.00)</td>
<td>4.68 (0.54)</td>
</tr>
<tr>
<td>English</td>
<td>Literacy</td>
<td>4.08 (0.67)</td>
<td>4.81 (0.38)</td>
</tr>
<tr>
<td></td>
<td>Speaking</td>
<td>3.89 (0.59)</td>
<td>4.71 (0.47)</td>
</tr>
</tbody>
</table>

Table 38: Mean percent of language ability by language by skill

Heritage speakers in the present study have spent more of their life in an English dominant environment, were educated in English and use more English in communicating with family and non-family than late bilinguals. This dominance in English likely impacts processing of the native language. Further analyses are needed to explore the exact effect of language use and exposure in Spanish-English bilinguals to determine if group differences are attributable to variations in how the two groups use language. Likewise, differences in proficiency ought to be further explored; however, measures of proficiency other than self-ratings are needed due to linguistic insecurity.

7.6 Future directions

The current study explored the subject-object relative clause processing asymmetry using subject- and object-relatives where the head noun was an object in the matrix clause. The majority of studies looking at relative clause processing asymmetry utilize stimuli where the head noun is a subject in the matrix clause. As previously mentioned, the processing asymmetry has been theorized to be a result of syntactic reanalysis (e.g., Clifton & Frazier, 1989), holding the filler in working memory (e.g., Wanner & Maratsos, 1978), and/or structural or thematic mismatch (Traxler, Morris, & Seely, 2002). In the present study, difficulty resulting from syntactic reanalysis or working memory demands are similar to previous studies using subject-embedded relative clauses, however, difficulty arising from structural or thematic mismatch are different for object-embedded relative clauses. The head noun in the current study was the matrix object which could have resulted in the subject-relative being harder to process due to the mismatch.
between the head noun’s thematic or structural role in the matrix clause and the subordinate clause. Thematic reanalysis elicits a posterior positivity 300-600ms post-event (Bornkessel et al., 2003). A future study is needed to explore whether the lack of the P375 and P420 is a result of heritage speakers being more sensitive to thematic mismatch. If there were greater positivity for object-subject relatives resulting from thematic mismatch and greater positivity for object-object relatives resulting from syntactic reanalysis or working memory demands, then no P400-like effects would be observed. To explore the effect of thematic mismatch in heritage speakers versus late bilinguals, a follow-up study is needed which includes subject- and object-relatives with a subject head noun.

8. Summary

In spite of being native speakers, heritage speakers process relative clauses differently than late bilinguals. While this study avoids explicit metalinguistic judgments of “correctness” or “grammaticality” in order to mitigate issues with linguistic insecurity, heritage speakers still appear to process their HL differently than individuals who acquired the L2 later. P300 and P600 effects were observed in late bilinguals and were absent in heritage speakers. The absence of late bilingual-like ERP effects in heritage speakers is attributed to greater processing difficulty for object-relatives and subject-relatives. This difficulty with the processing of complex syntactic structures is likely rooted in heritage speakers’ dominance in the ML.
The present chapter will report the results of another study of relative clause processing in heritage speakers and late bilinguals using pupillometry. Pupillometry studies utilize pupil diameter changes, or task-evoked pupillary responses (TEPR), as an index of cognitive load in response to some task. Section one of the chapter reviews the literature of research utilizing pupillometry in the study of language processing, the subject-object relative clause processing asymmetry, and multilingual language processing. Section two presents the research questions and hypotheses. Section three explains the methodology used in the present study. Section four explores the unprocessed pupil diameter data gathered. Section five reviews the data analyses and section six presents the results of the study. Section seven discusses the findings from section six. Lastly, section eight concludes.

1. Pupillometry and language processing

Pupillometry has increasingly been used in language research as a measure of cognitive load in processing and planning in a number of different domains. There has been an exponential increase in the language research literature of studies that use pupillometry in recent years (see Schmidtke (2017) for a review of 26 language studies using pupillometry from 1978-2016). The methodology has been shown to be sensitive to grammatical violations, complexity, and difficulty. Studies have investigated phenomena in auditory processing, written language processing, speech production and planning ranging from phoneme level processing (e.g., Tamasi et al., 2016) to pragmatics (e.g., Tromp, Hagoort, & Meyer, 2016) on participants as young as 3 months old (e.g., Hochmann and Papeo, 2014) to older adults in their 80s (e.g., Piquado, Isaacowitz, & Wingfield, 2010) in monolinguals and bilinguals.

1.1 Pupillometric studies of language

Processing costs have also been observed at the level of the lexicon in a number of studies. Unknown words (Ledoux et al., 2016), words that are new in the experiment (Võ et al., 2008), and words that do not match a given picture (Kuipers & Thierry, 2011; Renner & Włodarczak, 2017) elicit greater pupil dilation.
Pupillometry has also been used to explore the effects of frequency on lexical access. Researchers have found that lower frequency words require greater cognitive effort to access (Papesh & Goldinger, 2012; Chapman & Hallowell, 2015; Haro et al., 2017), consistent with previous research using other methodologies. Pupillary responses have also been shown to be sensitive to mispronunciation (Tamási et al., 2016) and mismatch of acoustic cues (Wagner, Toffanin, & Başkent, 2016).

In studies examining syntactic planning and production, greater pupillary responses were observed when recalling sentences of greater syntactic complexity (Stanners, Headley, & Clark, 1972) and when mentally transforming complex sentences that required ambiguity resolution (Schluroff et al., 1986). Sevilla, Maldonado, and Shalóm (2014) demonstrated that greater cognitive effort is needed in the planning and production of sentences that deviate from the canonical word/thematic order in Spanish. They observed greater pupil dilation when planning clitic-left dislocated object-verb-subject (OVS) sentences and passive construction sentences. In a between-group study, Sauppe (2017) observed that the planning costs of a non-canonical structure such as passive-voice sentences, depend on the syntax of the language. In comparing pupil diameter differences between active and passive sentences in Tagalog and German, they found that in a system such as Tagalog where passive-voice construction is similar to active-voice construction (the difference lies in inflectional morphology), planning effort is similar. Whereas in German, where the passive-voice construction is very different from active-voice construction, greater cognitive effort is required at various points in the sentence.

Researchers have also found that pupillary responses are modulated by syntactic complexity and ambiguity in sentence comprehension. Longer sentences (Lam, Youssef, & Clark, 2017), syntactically more complex sentences (Schluroff, 1982), and sentences with non-canonical word order such as object-verb-subject (OVS) in Danish (Wendt, Dau, & Hjortkjær, 2016) require more cognitive effort to process and elicit greater pupil dilation. Increased processing load is required and greater pupil dilation is observed when resolving temporary ambiguities that require structural reanalysis in the case of garden path sentences (Engelhardt, Ferreira, & Patsenko, 2010; Niikuni et al., 2015) and resolving the referent of a pronoun (Vogelzang, Hendriks, & van Rijn, 2016). Increased pupil diameter has also been observed in structures with filler-gap dependencies that require greater processing load to hold the unassociated filler
in memory (Just & Carpenter, 1993; Fernandez, 2013) and when processing grammatical gender violations (Demberg, Kiagia, & Sayeed, 2013; replicated in Demberg & Sayeed, 2016).

Additionally, pupillary responses have been found to index the increased processing demands associated with listening to degraded speech (Zekveld, Kramer, & Festen, 2010; Wendt, Dau, & Hjortkjær, 2016; Wagner, Toffanin, & Başkent, 2016), processing new information (Zellin et al., 2011), processing unexpected information (Ospina, 2016), making additional pragmatic inferences (Tromp, Hagoort, & Meyer, 2016), and retaining sentences in memory (Wright & Kahneman, 1971). Pupillary responses have also been observed when processing illicit events such as semantic violations (Demberg, Kiagia, & Sayeed, 2013; Demberg & Sayeed, 2016) and prosodic violations (Engelhardt, Ferreira, & Patsenko, 2010; Zellin et al., 2011).

1.2 Pupillometric studies of relative clause processing

Only a handful of studies have looked at the subject-object relative clause processing asymmetry as indexed by pupillary responses. These studies have examined the processing asymmetry in German and English utilizing both written and aurally-presented stimuli in adult monolingual speakers. They all found more or greater pupillary responses to object-relatives compared to subject-relatives.

Subject-Relatives (RCSS)
(27) The reporter [that ____ attacked the senator] admitted the error

Object-Relatives (RCSO)
(28) The reporter [that the senator attacked ____] admitted the error

Just and Carpenter (1993) looked at the processing of English relative clauses, such as in (27) and (28), during a reading task. They found that object-relatives elicited greater mean dilation than subject-relatives which they argued reflected intensity of thought or processing. They further grouped participants in high, medium, and low working memory groups as measured by a reading span task. Mean pupil dilation across condition trended towards being greater for individuals with lower working memory than higher working memory, although these differences were not significant. They argued that individuals with lower working memory span had more limited resources and were therefore utilizing a higher proportion of their
available resources for processing subject-relatives in addition to object-relatives than individuals with higher working memory.

The findings of Just and Carpenter (1993) were conceptually replicated and extended by Piquado, Isaacowitz, and Wingfield (2010). They included subject- and object-relatives, as in (27) and (28), but also included more complex subject- and object-relatives, as in (29) and (30), which contained adjectival modifiers on each of the noun phrases in the sentences. They auditorily presented these English relative clauses to younger and older adult participants and measured their pupillary responses during a two second period where participants were asked to retain the stimuli for recall. They additionally controlled for changes in pupillary reflexes as a result of senescence. Young adults exhibited the subject-object relative clause processing asymmetry (greater dilation for object-relatives than subject-relatives) in both the simple and complex conditions. However, in older adults, the processing asymmetry was only observed in the complex condition. Pupil dilation was comparable for object- and subject-relatives in the simple condition. The authors argued that this reflects more efficient processing on the part of older adults given their greater experience with the language and therefore object-relative clauses.

Subject-Relatives (RCSS)
(29) The professional gambler [that ___ signaled the suspicious dealer ] revealed the perfect card

Object-Relatives (RCSO)
(30) The professional gambler [that the suspicious dealer signaled ___ ] revealed the perfect card

The subject-object relative clause asymmetry has also been observed and replicated a number of times in German in both reading and listening tasks. The German relative clauses used in these experiments, as in (31) and (32), are ambiguous until the subordinate clause final, auxiliary verb is reached. Object-relatives are argued to be more difficult to process since they require reanalysis upon encountering the disambiguating auxiliary. In a dual-task experiment, a driving simulation task and a language comprehension task, with auditory presentation of the relative clauses, greater task-evoked pupillary responses were observed for object-relatives than subject-relatives (Engonopoulos, 2012; Engonopoulos, Sayeed, & Demberg, 2013; Demberg, 2013). Object-relatives elicited greater dilation as well as greater Index of Cognitive Activity (ICA: Marshall, 2002; et seq.) that Demberg and Sayeed (2016) explain as the
hyperbolic tangent-transformed number of rapid, small pupil dilations. The proprietary ICA is argued to be a better measure of cognition related task-evoked pupillary response in experiments where participants are free to move their gaze and the luminescence of the environment changes as it controls for pupillary reflexes related to light and eye movement. These dual-task findings were later replicated by Demberg and Sayeed (2016). In a single task studies, object-relatives elicited greater dilation and ICA during self-paced reading tasks in Demberg, Kiagia, and Sayeed (2013) and also replicated in Demberg and Sayeed (2016).

Subject-Relatives (RCSS)
(31) Die Nachbarin, [die einige der Mieter auf Schadensersatz verklagt hat ], traf sich gestern mit Angelika
The neighbor, who sued some of the tenants for damages, met Angelika yesterday

Object-Relatives (RCSO)
(32) Die Nachbarin, [die einige der Mieter auf Schadensersatz verklagt haben ], traf sich gestern mit Angelika
The neighbor, who some of the tenants sued for damages, met Angelika yesterday

While the subject-object relative clause processing asymmetry is well established as measurable using pupillometry, the findings of Just and Carpenter (1993) with individuals with high versus low working memory and of Piquado, Isaacowitz, and Wingfield (2010) with older versus younger adults, suggest that pupillometry is not an absolute measure of difficulty inherent in processing a structure like relative clauses, but impacted by the cognitive capacity and experience of an individual.

1.3 Pupillometric studies of multilingual processing

Relatively few studies have explored multilingual processing using pupillometry but have all found that bilinguals process language differently than monolinguals. In lexical processing, bilinguals were found to
be more sensitive to mispronounced words than monolinguals (Tamási et al., 2016). Differences in task-evoked pupillary response (TEPR) amplitude and latency were also found in syntactic processing of structures requiring syntactic reanalysis, ungrammaticality, agreement (Gutiérrez, 2013), and syntactic movement (Fernandez et al., 2017). Bilinguals were found to have decreased TEPR amplitude and increased latencies in response to syntactic processing effects. Difference has also been observed as a global measure of increased cognitive effort in listening to the non-native language. Non-native speakers had greater pupil size when listening to the L2 compared to monolingual speakers (Borghini & Hazan, 2018).

Previous research has shown that these differences are impacted by dominance, age of acquisition, and proficiency. In the processing of ungrammatical sentences due to pronoun agreement and of ungrammatical sentences with filler-gap dependencies, simultaneous bilinguals pattern similarly to monolinguals and differently from sequential bilinguals (Gutiérrez, 2013) likely reflecting differences related to dominance and age of acquisition. The effect of proficiency and age of acquisition has been found to have a continuous effect on pupillary response. Differences in pupil size and maximum dilation latency between bilinguals and monolinguals were minimized with greater L2 proficiency and early L2 age of acquisition (Schmidtke, 2014). Bilinguals’ pupillary responses are also impacted by the language that is being processed. Increased cognitive effort as indexed by greater pupillary responses has been observed in bilinguals when processing language that code-switched between the dominant to non-dominant language (Byers-Heinlein, Morin-Lessard, & Lew-Williams, 2017) and when bilinguals were translating and repeating words in the non-native language (Hyönä, Tommola, & Alaja, 1995).

Bilinguals have also been shown to be sensitive to similarities between the two linguistic systems they command. In domains where the two systems are similar, such as in the lexicon or syntax, less of a processing load is observed. Greater cognitive effort is observed in lexical access for non-cognate words compared to words that are identical cognates in the two languages (Guasch, Ferré, & Haro, 2017). Likewise, in simultaneous translation between languages, greater processing effort is seen when translating a verb-final structure into a language which only allows for verb-initial structures (Seeber & Kerzel, 2012).
2. Research question & hypothesis

The present study addresses the paucity of literature on pupillometric studies of bilingual processing. Previous studies of the subject-object relative clause asymmetry using task-evoked pupillary responses have exclusively focused on monolingual populations. We aim to extend this research by looking at the processing of relative clauses in bilinguals: late bilinguals and heritage speakers. Like ERPs, discussed in the previous chapter, the pupillometry experimental paradigm is well suited as a methodologically for heritage speaker bilingual studies as the effects of linguistic insecurity are mitigated by not requiring overt metalinguistic judgments. In this chapter, we ask the following two research questions:

1. Are task-evoked pupillary responses which indicate processing difficulty greater for object-relatives than subject-relatives in late bilinguals?

2. Do heritage speakers process object-relatives and subject-relatives differently when compared to late bilinguals when hearing Spanish sentences?

The subject-object asymmetry has been observed across a number of experimental methodologies and languages. We expect that the increased processing costs associated with object-relatives should be observable as differences in pupillary responses. We expect that object-relatives, being more difficult to process, should elicit greater pupil dilation than subject-relatives.

Our second research question is whether heritage speaker bilinguals process relative clauses similarly to late bilinguals. This will be assessed by comparing pupillary responses by bilingual group. Heritage speaker bilinguals’ command, use, and low confidence in their first-learnt language, Spanish, has been noted in the literature and their underperformance on metalinguistic tasks has been documented. These effects might impact pupillary responses. Alternatively, relative clauses in the first learnt language, Spanish, are structurally similar to relative clauses in English. Therefore, assuming positive transfer from the dominant language, we might hypothesize that heritage speakers will process relative clauses similar to late bilinguals. If heritage speaker bilinguals process relative clauses in a similar fashion as late bilinguals, we would expect that pupil dilation for heritage speaker bilinguals would not to be significantly different than for late bilinguals.
However, while heritage speaker bilinguals are indeed fluent speakers of Spanish, they are characteristically dominant in their second learnt language, English. Therefore, if we do observe differences in pupillary responses between the bilingual groups, this might indicate that language processing is affected by factors contributing to dominance, such as use and proficiency. If heritage speaker bilinguals process relative clauses in a significantly different way than late bilinguals, this should be reflected in their pupil dilation. For example, heritage speakers’ pupil dilation responses could be significantly smaller than the pupil dilation observed in late bilinguals.

3. Methods

3.1 Participants

Forty-three Spanish-English bilingual adults living in the New York City participated in the study at the CUNY Graduate Center in midtown Manhattan, NY and were compensated financially for their participation. The study protocol was approved by the CUNY institutional review board and written consent was obtained.

The participants were fluent Spanish and English speakers, have normal or corrected-to-normal vision, have normal hearing, have no history of a neurological disorder, and had not taken antihistamines the day of the pupillometry experiment.

Native Spanish-speaking participants were categorized as either heritage Spanish speakers (n=26) or late bilinguals (n=17) based on criteria commonly used in heritage speaker studies (Benmamoun, Montrul, & Polinsky, 2013) and the pre-determined inclusion criteria. Nearly three-quarters of the heritage speakers were born in the Anglophone US (65.38%, n=17) and the rest moved to the anglophone US before age 8 (34.62%, n=9, M=4.00, SD=2.00). Heritage speakers were raised speaking primarily Spanish until at least age 10 by Spanish-speaking immigrant parents originally from a Spanish-dominant country/region. Late bilinguals were born in a Spanish-dominant country/region and moved to the anglophone US at the age of 17 or older (M=25.29, SD=4.37).
3.2 Ancillary Measures

The pupil diameter recordings were made in the first of two test sessions. The second session 10-14 days later consisted of an acceptability judgment task, two n-back tasks (in English and Spanish), two n-back tasks (in English and Spanish), two speaking fluency tasks (in English and Spanish), and a Spanish syntax comprehension task. Demographic and language background was elicited in a 38-item questionnaire that was administered during the recruitment process. Results from the acceptability judgement task are not reported here.

3.2.1 Working memory

As pupillometry is a measure of processing load and differences in working memory between participants are a possible confound, we developed and administered two working memory tasks in Spanish: n-back, serial recall. Both tasks were administered in English and Spanish. The tasks were administered in the participant’s dominant language first. The tests were developed based on English tests used in Sprouse, Wagners, and Phillips’ (2012) study of working memory and syntactic island effects.

The n-back task had three levels of difficulty: two-back, three-back, and four-back. The tasks were administered using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) and a short break was given between tasks. The stimuli consisted of pseudo-randomized lists of eight visually-presented letters that were common between the Latin American Spanish and American English orthographies. D-prime scores, a measure of discrimination, were calculated for each test using the neuropsychology::dprime() function (Makowski, 2016). d' scores from the three-back are used in this study. Heritage speakers ($M=1.77, SD=0.75$) did not perform significantly different from late bilinguals ($M=1.53, SD=0.48$), $t(30)=-0.94, p=.36, r=.17$.

The serial recall task consisted of ten pseudorandomized lists of eight aurally-presented disyllabic Spanish words that were matched for orthographic and phonetic form (CVCVC), approximate log frequency, neighborhood density, and phonotactic probability. Participants were instructed to repeat “el” the to themselves during presentation to suppress articulatory repetition of the list and then given 30 seconds after a 500ms break to recall the list after presentation using a pencil on paper. The proportion of correct responses by position was scored and then summed. Accuracy scores could range from 0-8.
Responses within a generalized Levenshtein (edit) distance (Levenshtein, 1966) of 2 using the `utils::adist()` function (R Core Team, 2017) were scored as correct. Heritage speakers ($M=5.06$, $SD=0.97$) were not significantly different on the Spanish serial recall task than late bilinguals ($M=5.00$, $SD=0.86$), $t(30)=-0.16$, $p=.87$, $r=.03$.

The two working memory control variables were not significantly correlated across group ($r(30)=.00008$, 95% CI [-.35, .35], $p=.99$) or by bilingual group: late bilinguals ($r(8)=.20$, 95% CI [-.49, .74], $p=.58$), heritage speakers ($r(20)=-.06$, 95% CI [-.47, .37], $p=.80$).

### 3.2.2 Spanish Comprehension

Spanish proficiency in heritage speakers, who are typically English dominant, and late bilinguals was tested using the RISLUS Multilingual Syntax Test (RMST) (Klein & Martohardjono, 2009) described in chapter 3, section 4.1. Late bilinguals performed at ceiling (Accuracy: $M=.95$, $SD=.03$, range: .91-.100) on the RMST. Similarly, heritage speakers performed at ceiling (Accuracy: $M=.95$, $SD=.04$, range: .86-.100) on the RMST, indicating fluency in Spanish. Differences in accuracy between groups was not tested with a t-test as both groups being at ceiling made the statistical which assumes a Gaussian distribution not appropriate. The test was integrated into the experimental protocol after data collection had begun. RMST data is only reported for a subset of late bilinguals (41.18%, $n=7$) and heritage speakers (69.23%, $n=18$), however, these results are consistent with the Spanish RMST results for heritage speakers in section 3.1.

### 3.2.3 Spanish Oral Fluency

Spanish oral fluency was also measured using a story-telling task (Berman & Slobin, 1994) following Polinsky (2008). Participants were asked to produce a story using one of two frog story elicitation books: *Frog, Where Are You?* (Mayer, 1969), *One Frog Too Many* (Mayer & Mayer, 1975). Story narratives were recorded using Audacity® 2.0.3 (Audacity Team, 2014) and transcribed with fillers and dysfluencies by a team of Spanish-speaking coders aided with Google’s Web Speech API Demonstration (https://www.google.com/intl/en/chrome/demos/speech.html) setting the Spanish variety to the appropriate Spanish region. Oral fluency was calculated as words per minute, including fillers and code-switched words, starting 30 seconds after the narrative beginning after. Late bilinguals had a faster
Spanish speech rate ($M=137.20$, $SD=21.91$) than heritage speakers ($M=118.05$, $SD=25.75$), although this difference only approached significance, $t(30)=2.04$, $p=.05$, $r=.35$. Heritage speakers and late bilinguals in this study were not reliably different from each other in terms of Spanish comprehension of complex sentences, working memory, and oral fluency.

3.3 Pupillometry

3.3.1 Stimuli
Stimuli were presented auditorily and were identical to stimuli used in the ERP task (see chapter 4, section 3.2 for details of stimuli).

3.3.2 Experimental design and procedures
All target stimuli in the two conditions ($n=30$) were grammatical and were interspersed with grammatical and ungrammatical sentences ($n=610$) presented in the same context sentence-target question pair format. These other sentences represent conditions reported in other studies and served as fillers for the present study. Trials from different conditions were evenly distributed over five blocks and pseudorandomized so that items from the same condition never appeared consecutively. Trials were presented using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). The pupillometry portion of the experiment took approximately 1.5 hours, including set-up, with breaks every 10 minutes.

Participants were seated in a padded, plastic chair, 70cm from a Tobii TX300 eye-tracker and external speakers in a dimly lit, shielded IAC booth and instructed to minimize eye movements and eye blinks. Instructions were aurally presented in English and Spanish before a short practice session. Participants were not asked to make any metalinguistic judgments about the target items in the pupillometry experiment. A fixation marker (a white cross + in the center of a black Tobii monitor screen) was provided throughout the auditory blocks. Following 40% of the auditorily-presented trials, the participants were prompted on screen to answer a written comprehension question about the sentences in the trial that was presented in white font on the black background. The comprehension question was unrelated to the research questions in this study and served the purpose of ensuring participants’ continued attention. The auditory block resumed after participants answered the question.
Pupil diameter and gaze location were recorded with Tobii TX300 infrared cameras for each eye separately. Data were recorded for the whole trial (including the context and target) and 1 second (1000ms) before and after the trial at 60Hz (one sample every 16.67 milliseconds).

3.3.1 Data processing
Pre-processing was performed by-item by-participant by-eye. Pupil diameter data were smoothed using a low-pass, 10Hz, second-order Butterworth filter, the signal::butter() function (signal developers, 2013), to remove high frequency noise due to participant movements, eye tremors, and non-spherical eye shape (McLaren et al., 1992, as cited in Klingner, Kumar, & Hanrahan, 2008). Removing artifacts was done to make subsequent interpolation of missing values more reliable (Jackson & Sirois, 2009). Missing samples from a single eye were imputed with a linear regression based on data from the other eye. Data from both eyes were averaged and samples that changed more than 0.50% in diameter from the previous sample were removed. Linear interpolation of missing values in the eye-average data were performed using the zoo::na.approx() function (Zeilis & Grothendieck, 2005). The pupil size recorded by the eyetracker is in arbitrary units and was not transformed by a scaling factor to represent actual pupil diameter in micrometers. Pupil size measures are corrected with a 200ms pre-stimulus onset (while fixating on a cross +) baseline to reflect change from baseline in arbitrary units.

4. Pupil diameter data
Grand mean waveforms by relative clause condition by bilingual group of untransformed dilation by millisecond are plotted in Figure 38 using the ggplot package (Wickham, 2009). Visual inspection of the grand mean waveforms by relative clause condition by bilingual group showed greater dilation was elicited for object-relatives than subject-relatives in late bilinguals, and the opposite effect was elicited in heritage speakers. For late bilinguals, object-relatives elicited dilation starting at the onset of the stimuli and continued until the end of the trial. For subject-relatives, the same pattern is seen with dilation being observed from the onset of the stimuli until the end of the sample, although with a less steep slope of dilation for object-relatives. For heritage speakers, both the subject-relatives and the object-relatives elicited dilation starting around 400ms post-stimuli onset. The object-relatives appear to peak around
1250ms and reached asymptote. Dilation for subject-relatives also follows the same trend of sharply increasing between 400ms-1250ms and then continues to dilate until the end of the sample with a less steep slope than observed in the earlier 400ms-1250ms window.

Figure 38: Grand mean waveforms for subject-relatives (dashed, blue line) and object-relatives (solid, red line) by-bilingual group by millisecond.

5. Data analysis

Pupil dilation data was further processed and four dependent measures of task-evoked pupillary responses (TEPR) were calculated, as illustrated in Figure 39. First, pupil size recordings were pre-processed to remove artifacts resulting from blinks, looking away from the screen, low validity data, and baseline correct trials as detailed in section 3.3.1 above. Then due to the leptokurtic distribution and the presence of outliers in the pupil diameter data, 95% winsorized, square root transformed data were used. After which, four dependent variables were derived from the normalized pupil size data. Lastly, the data were fit with linear mixed-effects models for hypothesis testing separately for each of the four derived TEPR variables.

Figure 39: General pupillometry data analysis flowchart.
5.1 Calculate TEPR

Various measures have been derived in research exploring task-evoked pupillary responses (TEPR) as a measure of cognitive load, such as dilation slope, percent change, latency to maxima, etc. In the present study, processing load was operationalized as four separate measures of mean pupil size using different sized bins: trial-average pupil diameter, maximum pupil diameter, epoch-average pupil diameter, and window-average pupil diameter.

5.1.1 Trial-average TEPR

Average of baseline corrected, transformed pupil diameter from stimulus onset to 1000ms stimulus offset was calculated by-item and by-participant. Trial-average pupil diameter is a global measure which reflects average cognitive effort of over the course of the whole trial controlling for different trial length. This derived TEPR variable is illustrated in Figure 40 showing a single trial waveform from one participant. The shaded purple area under the curve represents the data that was included in the average.

5.1.2 Maximum TEPR

The maximum average value from among 100ms bin averages of baseline corrected, transformed pupil diameter from stimulus onset to 1000ms stimulus offset was calculated by-item and by-participant. Maximum pupil diameter reflects maximum cognitive effort using the average of a bin to control for high frequency noise in the data. This derived TEPR variable is illustrated in Figure 41 showing a single trial waveform for one participant. The green shaded area indicates the data included in the maximum TEPR calculation.

Figure 40: Single trial waveform for one participant. Purple shaded area indicating data included in calculation of trial-average TEPR.

Figure 41: Single trial waveform for one participant. Green shaded area indicating data included in calculation of maximum TEPR.
waveform from one participant. The green bar represented data that was included in calculating the bin average.

5.1.3 **Epoch-average TEPR**

Average of baseline corrected, transformed pupil diameter from onset of the subordinate verb to 1000ms stimulus offset was calculated by-item and by-participant. Epoch-average pupil diameter is a measure that reflects average cognitive effort directly in response to the first point of structural difference in subject- and object-relative clauses not including the lead-up to the gap/subordinate subject controlling for different trial length. This derived TEPR variable is illustrated in Figure 42 showing a single trial waveform from one participant. The shaded salmon area under the curve represents the data that was included in the average.

5.1.4 **Window-average TEPR**

A 500ms window average of baseline corrected, transformed pupil diameter from stimulus onset to 1000ms stimulus offset was calculated by-item and by-participant. Window-average pupil diameter is a measure which reflects average cognitive effort for 500ms bin. Data are analyzed by bin to determine roughly the point at which processing of the two structure diverges controlling for high frequency noise. This derived TEPR variable is illustrated in Figure 43 showing a single trial waveform from one participant. The pink dots represent the window-average values from the 500ms bins for the whole trial.
5.2 Linear mixed-effects modeling

Hypothesis testing was done using linear mixed-effects models as in the ERP study. Specific modeling details are in chapter 4, section 5.2. All mixed-effects models for the TEPR derived dependent variables included working memory control variables (Spanish serial recall accuracy, 3-back d’ scores) as fixed effects. It should be further noted that when interpreting the coefficients of the models, estimates, and the predicted marginal means are calculated assuming working memory as zero.

5.3 Across TEPR analysis

Correlations by- and across- bilingual group of participant-average TEPR values are run to explore the relationship between the different derived TEPR variables found in the study using the stats::cor.test() function in R (R Core Team, 2017). TEPR difference value was calculated by subtracting mean subject-relative TEPR value from mean object-relative TEPR value by participant. The Pearson’s r correlation coefficient, p-value, and 95% confidence intervals of the r coefficient are compared across and within bilingual groups.

6. Results

Results of task-evoked pupillary responses indexing processing load are reported by TERP variable: trial-average TEPR, maximum TEPR, epoch-average TEPR, window-average TEPR by window. A table of condition and bilingual group means is given in each section. The most parsimonious model for each TEPR variable is reported below and interpreted.

6.1 Average TEPR – Whole trial

Object-relatives elicited greater trial-average dilation than subject-relatives for both groups. Mean trial-average dilation by bilingual group and condition is plotted with 95% confidence interval error bars in Figure 44. Mean trial-average dilation by bilingual group and condition is reported in Table 39.
**Figure 44:** Mean trial-average dilation by bilingual group and condition with 95% confidence interval error bars.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-Relatives</td>
<td>Late Bilinguals</td>
<td>3.48 (8.14)</td>
<td>5.14 (8.77)</td>
<td>4.62 (8.60)</td>
</tr>
<tr>
<td></td>
<td>Heritage Speakers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object-Relatives</td>
<td>Late Bilinguals</td>
<td>6.03 (8.46)</td>
<td>5.35 (8.54)</td>
<td>5.56 (8.51)</td>
</tr>
<tr>
<td></td>
<td>Heritage Speakers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Late Bilinguals</td>
<td>4.75 (8.38)</td>
<td>5.24 (8.65)</td>
<td>5.09 (8.57)</td>
</tr>
</tbody>
</table>

**Table 39:** Mean trial-average dilation by bilingual group and condition. (Standard deviations in parentheses).

Across group, trial-average dilation for object-relatives ($M$=5.56, $SD$=8.51) was greater than subject-relatives ($M$=4.62, $SD$=8.60) indicating that across group, object-relatives were associated with greater average cognitive effort over the whole trial. Heritage speakers appeared to experience greater difficulty with all relative clauses as heritage speakers ($M$=5.24, $SD$=8.65) had greater trial-average dilation than late bilinguals ($M$=4.75, $SD$=8.38). Both late bilinguals and heritage speakers exhibited the subject-object processing asymmetry as indexed by trial-average dilation, although the difference between conditions was greater for late bilinguals than heritage speakers. Late bilinguals had greater trial-average dilation for object-relatives ($M$=6.03, $SD$=8.46) than subject-relatives ($M$=3.48, $SD$=8.14). Heritage speakers also had greater trial-average dilation for object-relatives ($M$=5.35, $SD$=8.54) than subject-relatives ($M$=5.14,
SD=8.77), although not of the same magnitude as late bilinguals. Heritage speakers had greater trial-average dilation for subject-relatives that late bilinguals had for subject-relatives.

Figure 45: Linear mixed-effects models for trial-average dilation. Significant parsimonious model in red type.

To determine whether differences between condition and bilingual group were statistically significant, trial-average dilation was modeled following the procedure outlined in section 5.3 and visualized in Figure 45. The maximal model, which was also the most parsimonious model, shown in Table 40, retained the predictor variables of group, condition, and the interaction, controlling for working memory, had a significantly better fit than the working memory only model ($\chi^2(3)=7.895, p=.048$) and a simpler model without the interaction ($\chi^2(1)=4.414, p=.036$).

The intercept of the model was not significant, meaning that late bilingual trial-average dilation for subject-relatives did not significantly differed from the mean across condition and across group, controlling for working memory ($M=5.09, SD=8.57$), $B=-2.936, SE(B)=3.689, t(33)=0.796, p=.432$. The predictor variable of group was not significant, meaning that although heritage speakers had more trial-average dilation for subject-relatives than late bilinguals did for subject-relatives, this difference was not significant, $B=1.751, SE(B)=1.402, t(45)=1.249, p=.218$. The predictor variable of condition was
significant, meaning that late bilinguals had significantly more trial-average dilation for object-relatives than subject-relatives, \( B=2.545, SE(B)=0.921, t(928)=2.763, p=.006 \). Lastly, the interaction between bilingual group and condition was significant, meaning that the difference between subject-relatives and object-relatives was significantly different between bilingual group, \( B=-2.337, SE(B)=1.111, t(928)=-2.103, p=.036 \). The working memory control variables as measured by the n-back \( (B=-0.445, SE(B)=0.894, t(32)=-0.498, p=.622) \) and Spanish serial recall \( (B=0.245, SE(B)=1.111, t(32)=0.381, p=.706) \) tasks, were not significant predictors of trial-average TEPR.

![Table 40](Image)

\( \text{Formula in R: Average Dilation ~ cond * gen + n.back + sr + (1 | part) + (1 | item)} \)

\( ^\dagger \ p<.1 \quad ^\prime p<.05 \quad ^\prime\prime p<.01 \quad ^\prime\prime\prime p<.001 \)

Table 40: Maximal linear mixed-effects model of trial-average dilation elicited by subject- and object-relative clauses.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.936</td>
<td>3.689</td>
<td>33</td>
<td>0.796</td>
</tr>
<tr>
<td>\text{Experimental} Group(HS)</td>
<td>1.751</td>
<td>1.402</td>
<td>45</td>
<td>1.249</td>
</tr>
<tr>
<td>RC(object)</td>
<td>2.545</td>
<td>0.921</td>
<td>928</td>
<td>2.763</td>
</tr>
<tr>
<td>Group:RC</td>
<td>-2.337</td>
<td>1.111</td>
<td>928</td>
<td>-2.103</td>
</tr>
<tr>
<td>\text{Control} n-back</td>
<td>-0.445</td>
<td>0.894</td>
<td>32</td>
<td>-0.498</td>
</tr>
<tr>
<td>Serial Recall</td>
<td>0.245</td>
<td>1.111</td>
<td>32</td>
<td>0.381</td>
</tr>
</tbody>
</table>

The predicted values from the maximal model were used to calculate descriptive statistics, shown in Table 41 and as outlined in section 5.3, to compare groups and explore the subject-object asymmetry by- and across- groups after having controlled for individual- and item-level variance and working memory.

Across condition, although heritage speakers \( (M_p=5.273, SE_p=0.713, 95\% CI_p [3.822, 6.724]) \) did have...
more trial-average dilation than late bilinguals ($M_p=4.690$, $SE_p=1.063$, 95% CI $[2.525, 6.855]$), this difference was not significant. Similarly, across group, object-relatives ($M_p=5.670$, $SE_p=0.694$, 95% CI $[4.273, 7.067]$) elicited more trial-average dilation than subject-relatives ($M_p=4.293$, $SE_p=0.694$, 95% CI $[2.896, 5.690]$), this difference was not significant. For late bilinguals, object-relatives ($M_p=5.963$, $SE_p=1.158$, 95% CI $[3.630, 8.295]$) elicited significantly more trial-average dilation than subject-relatives ($M_p=3.418$, $SE_p=1.158$, 95% CI $[1.085, 5.750]$). For heritage speakers, although object-relatives ($M_p=5.377$, $SE_p=0.777$, 95% CI $[3.812, 6.942]$) elicited slightly more trial-average dilation than subject-relatives ($M_p=5.169$, $SE_p=0.777$, 95% CI $[3.603, 6.734]$), this difference was not significant. Lastly, although heritage speakers had more trial-average dilation for subject-relatives that late bilinguals did for subject-relatives, this difference was not significant.

6.2 Maximum TEPR

Object-relatives elicited greater maximum TEPR amplitude than subject-relatives for both groups. Mean maximum dilation by bilingual group and condition is plotted with 95% confidence interval error bars in Figure 46. Mean maximum dilation by bilingual group and condition is reported in Table 42.

![Figure 46: Mean maximum dilation by bilingual group and condition with 95% confidence interval error bars.](image)
Across group, maximum dilation for object-relatives ($M=13.72$, $SD=6.38$) was greater than subject-relatives ($M=12.45$, $SD=6.67$) indicating that across group, object-relatives were associated with greater cognitive effort at one point in time. Heritage speakers appeared to experience greater difficulty with all relative clause items as heritage speakers ($M=13.16$, $SD=6.50$) had greater maximum dilation than late bilinguals ($M=12.92$, $SD=6.68$). Both late bilinguals and heritage speakers exhibited the subject-object processing asymmetry as indexed by maximum dilation, although the difference between conditions was greater for late bilinguals than heritage speakers. Late bilinguals had greater maximum dilation for object-relatives ($M=13.53$, $SD=6.26$) than subject-relatives ($M=12.79$, $SD=6.71$). Heritage speakers also had greater maximum dilation for object-relatives ($M=14.14$, $SD=6.64$) than subject-relatives ($M=11.70$, $SD=6.52$), although not of the same magnitude as late bilinguals. Heritage speakers had greater maximum dilation for subject-relatives that late bilinguals had for subject-relatives.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group</th>
<th>Late Bilinguals</th>
<th>Heritage Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-Relatives</td>
<td>11.70</td>
<td>12.79</td>
<td>12.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.52)</td>
<td>(6.71)</td>
<td>(6.67)</td>
<td></td>
</tr>
<tr>
<td>Object-Relatives</td>
<td>14.14</td>
<td>13.53</td>
<td>13.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.64)</td>
<td>(6.26)</td>
<td>(6.38)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>12.92</td>
<td>13.16</td>
<td>13.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.68)</td>
<td>(6.50)</td>
<td>(6.55)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 42: Mean trial maximum dilation by bilingual group and condition. (Standard deviations in parentheses).**

To determine whether differences between condition and bilingual group were statistically significant, maximum dilation was modeled following the procedure outlined in section 5.3 and visualized in Figure 47. The maximal model, which was also the most parsimonious model, shown in Table 43, retained the predictor variables of group, condition, and the interaction, controlling for working memory, had a significantly better fit than the working memory only model ($\chi^2(3)=12.944$, $p=.005$) and a simpler model without the interaction ($\chi^2(1)=4.276$, $p=.039$).

The intercept of the model was significant, meaning that late bilingual maximum dilation for subject-relatives significantly differed from the mean across condition and across group, controlling for working memory ($M=13.09$, $SD=6.55$), $B=13.733$, $SE(B)=3.132$, $t(32.8)=4.386$, $p<.001$. The predictor
Figure 47: Linear mixed-effects models for maximum dilation. Significant parsimonious model in red type.

variable of group was not significant, meaning that although heritage speakers had greater maximum dilation for subject-relatives than late bilinguals did for subject-relatives, this difference was not significant, $B=1.323$, $SE(B)=1.170$, $t(41.7)=1.131$, $p=.264$. The predictor variable of condition was significant, meaning that late bilinguals had significantly greater maximum dilation for object-relatives than subject-relatives, $B=2.443$, $SE(B)=0.696$, $t(227.9)=3.508$, $p=.001$. Lastly, the interaction between bilingual group and condition was significant, meaning that the difference between subject-relatives and object-relatives was significantly different between bilingual group, $B=-1.706$, $SE(B)=0.824$, $t(898.7)=-2.070$, $p=.039$. The working memory control variables as measured by the n-back ($B=-0.912$, $SE(B)=0.760$, $t(32.0)=-1.200$, $p=.239$) and Spanish serial recall ($B=-0.128$, $SE(B)=0.548$, $t(32.0)=-0.233$, $p=.817$) tasks, were not significant predictors of maximum TEPR amplitude.

The predicted values from the maximal model were used to calculate descriptive statistics, shown in Table 44 and as outlined in section 5.3, to compare groups and explore the subject-object asymmetry by- and across- groups after having controlled for individual- and item-level variance and working memory. Across condition, although heritage speakers ($M_p=13.233$, $SE_p=0.609$, 95% $CI_p$ [11.922, 14.474]) did have more maximum dilation than late bilinguals ($M_p=12.763$, $SE_p=0.906$, 95% $CI_p$ [10.918,
this difference was not significant. Across group, object-relatives ($M_p=13.793$, $SE_p=0.586$, 95% CI [12.609, 14.977]) elicited significantly more maximum dilation than subject-relatives ($M_p=12.203$, $SE_p=0.586$, 95% CI [11.019, 13.387]). For late bilinguals, object-relatives ($M_p=13.984$, $SE_p=0.971$, 95% CI [12.026, 15.943]) elicited significantly more maximum dilation than subject-relatives ($M_p=12.203$, $SE_p=0.971$, 95% CI [11.019, 13.387]). For heritage speakers, although object-relatives ($M_p=13.601$, $SE_p=0.655$, 95% CI [12.280, 14.923]) elicited slightly more maximum dilation than subject-relatives ($M_p=12.865$, $SE_p=0.655$, 95% CI [11.543, 14.186]), this difference was not significant. Lastly, although heritage speakers had more maximum dilation for subject-relatives that late bilinguals did for subject-relatives, this difference was not significant.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>13.733</td>
<td>3.132</td>
<td>32.8</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Experimental</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group(HS)</td>
<td>1.323</td>
<td>1.170</td>
<td>41.7</td>
<td>.264</td>
</tr>
<tr>
<td>RC(object)</td>
<td>2.443</td>
<td>0.696</td>
<td>227.9</td>
<td>.001 ***</td>
</tr>
<tr>
<td>Group:RC</td>
<td>-1.706</td>
<td>0.824</td>
<td>898.7</td>
<td>.039 *</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n-back</td>
<td>-0.912</td>
<td>0.760</td>
<td>32.0</td>
<td>.239</td>
</tr>
<tr>
<td>Serial Recall</td>
<td>-0.128</td>
<td>0.548</td>
<td>32.0</td>
<td>.817</td>
</tr>
</tbody>
</table>

Formula in R: Maximum Dilation ~ cond * gen + n.back + sr + (1 | part) + (1 | item)

Table 43: Maximal linear mixed-effects model of trial maximum dilation elicited by subject- and object-relative clauses.

<table>
<thead>
<tr>
<th>Group</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>$C_{lower}$</th>
<th>$C_{upper}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Late Bilingual</td>
<td>12.763</td>
<td>0.906</td>
<td>32.24</td>
<td>10.918</td>
<td>14.608</td>
</tr>
<tr>
<td>Heritage Speaker</td>
<td>13.233</td>
<td>0.609</td>
<td>32.30</td>
<td>11.992</td>
<td>14.474</td>
</tr>
<tr>
<td>Relative Clause</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>12.203</td>
<td>0.586</td>
<td>40.63</td>
<td>11.019</td>
<td>13.387 *</td>
</tr>
<tr>
<td>Object</td>
<td>13.793</td>
<td>0.586</td>
<td>40.63</td>
<td>12.609</td>
<td>14.977</td>
</tr>
<tr>
<td>Late Bilingual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject-relatives</td>
<td>11.541</td>
<td>0.971</td>
<td>42.10</td>
<td>9.583</td>
<td>13.500 *</td>
</tr>
<tr>
<td>Object-relatives</td>
<td>13.984</td>
<td>0.971</td>
<td>42.10</td>
<td>12.026</td>
<td>15.943</td>
</tr>
<tr>
<td>Heritage Speaker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject-relatives</td>
<td>12.865</td>
<td>0.655</td>
<td>41.38</td>
<td>11.543</td>
<td>14.186 n.s.</td>
</tr>
<tr>
<td>Object-relatives</td>
<td>13.601</td>
<td>0.655</td>
<td>41.38</td>
<td>12.280</td>
<td>14.923</td>
</tr>
</tbody>
</table>

Table 44: Predicted marginal means from maximal model of trial maximum dilation elicited by subject- and object-relative clauses.
6.3 Average TEPR – Epoch

Object-relatives elicited greater epoch-average dilation than subject-relatives for both groups. Mean epoch-average dilation by bilingual group and condition is plotted with 95% confidence interval error bars in Figure 48. Mean epoch-average dilation by bilingual group and condition is reported in Table 45.

![Figure 48: Mean epoch-average dilation by bilingual group and condition with 95% confidence interval error bars.](image)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Late Bilinguals</th>
<th>Heritage Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-Relatives</td>
<td>5.03 (9.91)</td>
<td>6.29 (10.84)</td>
<td>5.90 (10.57)</td>
</tr>
<tr>
<td>Object-Relatives</td>
<td>8.89 (9.99)</td>
<td>6.34 (11.62)</td>
<td>7.14 (11.19)</td>
</tr>
<tr>
<td>Total</td>
<td>6.96 (10.12)</td>
<td>6.32 (11.23)</td>
<td>6.52 (10.89)</td>
</tr>
</tbody>
</table>

*Table 45: Mean epoch-average dilation by bilingual group and condition. (Standard deviations in parentheses).*

Across group, epoch-average dilation for object-relatives ($M=7.14$, $SD=11.19$) was greater than subject-relatives ($M=5.90$, $SD=10.57$) indicating that across group, object-relatives were associated with greater average cognitive effort over the whole epoch after the onset of the subordinate verb. Late bilinguals appeared to experience slightly greater difficulty with all relative clause items as late bilinguals ($M=6.96$, $SD=10.89$).
had greater epoch-average dilation than heritage speakers ($M=6.32$, $SD=11.23$). Both late bilinguals and heritage speakers exhibited the subject-object processing asymmetry as indexed by epoch-average dilation, although the difference between conditions was far greater for late bilinguals than heritage speakers. Late bilinguals had greater epoch-average dilation for object-relatives ($M=8.89$, $SD=9.99$) than subject-relatives ($M=5.03$, $SD=9.91$). Heritage speakers also had greater epoch-average dilation for object-relatives ($M=6.34$, $SD=11.62$) than subject-relatives ($M=6.29$, $SD=10.84$), although this difference was very small. Heritage speakers had slightly greater epoch-average dilation for subject-relatives that late bilinguals had for subject-relatives.

Figure 49: Linear mixed-effects models for epoch-average dilation. Significant parsimonious model in red type.

To determine whether differences between condition and bilingual group were statistically significant, epoch-average dilation was modeled following the procedure outlined in section 5.3 and visualized in Figure 49. The maximal model, which was also the most parsimonious model, shown in Table 46, retained the predictor variables of group, condition, and the interaction, controlling for working memory, had a significantly better fit than the working memory only model ($\chi^2(3)=10.848$, $p=.013$) and a simpler model without the interaction ($\chi^2(1)=7.236$, $p=.007$).
The intercept of the model was not significant, meaning that late bilingual epoch-average dilation for subject-relatives did not significantly differ from the mean across condition and across group, controlling for working memory ($M=6.52$, $SD=10.89$), $B=4.276$, $SE(B)=4.565$, $t(33.1)=0.937$, $p=.356$. The predictor variable of group was not significant, meaning that although heritage speakers had more epoch-average dilation for subject-relatives than late bilinguals did for subject-relatives, this difference was not significant, $B=1.503$, $SE(B)=1.743$, $t(45.8)=0.863$, $p=.393$. The predictor variable of condition was significant, meaning that late bilinguals had significantly more epoch-average dilation for object-relatives than subject-relatives, $B=3.861$, $SE(B)=1.172$, $t(928)=3.294$, $p=.001$. Lastly, the interaction between bilingual group and condition was significant, meaning that the difference between subject-relatives and object-relatives was significantly different between bilingual group, $B=-3.810$, $SE(B)=1.414$, $t(928)=-2.695$, $p=.007$. The working memory control variables as measured by the n-back ($B=-1.116$, $SE(B)=1.106$, $t(32.0)=-1.009$, $p=.320$) and Spanish serial recall ($B=0.493$, $SE(B)=0.797$, $t(32.0)=0.618$, $p=.541$) tasks, were not significant predictors of epoch-average TEPR.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.276</td>
<td>4.565</td>
<td>33.1</td>
<td>0.937</td>
</tr>
<tr>
<td><strong>Experimental</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group(HS)</td>
<td>1.503</td>
<td>1.743</td>
<td>45.8</td>
<td>0.863</td>
</tr>
<tr>
<td>RC(object)</td>
<td>3.861</td>
<td>1.172</td>
<td>928</td>
<td>3.294</td>
</tr>
<tr>
<td>Group:RC</td>
<td>-3.810</td>
<td>1.414</td>
<td>928</td>
<td>-2.695</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n-back</td>
<td>-1.116</td>
<td>1.106</td>
<td>32.0</td>
<td>-1.009</td>
</tr>
<tr>
<td>Serial Recall</td>
<td>0.493</td>
<td>0.797</td>
<td>32.0</td>
<td>0.618</td>
</tr>
</tbody>
</table>

**Formula in R:** Average Dilation $\sim$ cond * gen + n.back + sr + (1 | part) + (1 | item)

† $p<.1$ †† $p<.05$ ††† $p<.01$ †††† $p<.001$

Table 46: Maximal linear mixed-effects model of epoch-average dilation elicited by subject- and object-relative clauses.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>C_{lower}</th>
<th>C_{upper}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late Bilingual</td>
<td>6.795</td>
<td>1.315</td>
<td>32.00</td>
<td>4.117</td>
</tr>
<tr>
<td>Heritage Speaker</td>
<td>6.393</td>
<td>0.881</td>
<td>32.00</td>
<td>4.597</td>
</tr>
<tr>
<td><strong>Relative Clause</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>5.616</td>
<td>0.862</td>
<td>46.17</td>
<td>3.881</td>
</tr>
<tr>
<td>Object</td>
<td>7.572</td>
<td>0.862</td>
<td>46.17</td>
<td>5.836</td>
</tr>
<tr>
<td><strong>Late Bilingual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject-relatives</td>
<td>4.864</td>
<td>1.439</td>
<td>45.92</td>
<td>1.967</td>
</tr>
<tr>
<td>Object-relatives</td>
<td>8.725</td>
<td>1.439</td>
<td>45.92</td>
<td>5.828</td>
</tr>
<tr>
<td><strong>Heritage Speaker</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject-relatives</td>
<td>6.368</td>
<td>0.966</td>
<td>46.09</td>
<td>4.423</td>
</tr>
<tr>
<td>Object-relatives</td>
<td>6.418</td>
<td>0.966</td>
<td>46.09</td>
<td>4.474</td>
</tr>
</tbody>
</table>

Table 47: Predicted marginal means from maximal model of epoch-average dilation elicited by subject- and object-relative clauses.
The predicted values from the maximal model were used to calculate descriptive statistics, shown in Table 47 and as outlined in section 5.3, to compare groups and explore the subject-object asymmetry by- and across- groups after having controlled for individual- and item-level variance and working memory. Across condition and an opposite trend than the other dependent variables, late bilinguals ($M_p=6.795$, $SE_p=1.315$, 95% CI $[4.117, 9.473]$) had more epoch-average dilation than heritage speakers ($M_p=6.393$, $SE_p=0.881$, 95% CI $[4.597, 8.188]$) although this difference was not significant. Across group, object-relatives ($M_p=7.572$, $SE_p=0.862$, 95% CI $[5.836, 9.307]$) elicited significantly more epoch-average dilation than subject-relatives ($M_p=5.616$, $SE_p=0.862$, 95% CI $[3.881, 7.351]$). For late bilinguals, object-relatives ($M_p=8.725$, $SE_p=1.439$, 95% CI $[5.828, 11.623]$) elicited significantly more epoch-average dilation than subject-relatives ($M_p=4.864$, $SE_p=1.439$, 95% CI $[1.967, 7.762]$). For heritage speakers, although object-relatives ($M_p=6.418$, $SE_p=0.966$, 95% CI $[4.474, 8.363]$) elicited slightly more epoch-average dilation than subject-relatives ($M_p=6.368$, $SE_p=0.966$, 95% CI $[4.423, 8.312]$), this difference was not significant. Lastly, although heritage speakers had more epoch-average dilation for subject-relatives that late bilinguals did for subject-relatives, this difference was not significant.

![Marginal Mean Dilation by Condition by Group](image)

**Figure 50:** Mean average dilation for 500ms windows by bilingual group and condition with 95% confidence interval error bars.

### 6.4 Average TEPR – 500ms window

Average TEPR by 500ms window from 0ms-4000ms was modeled separately. Fit of each maximal model was compared to the null model. Significance of each window model is reported in Table 48 with the chi-
square statistic and $p$-value of the fit of each maximal model. Models in window 3 1001-1500ms ($\chi^2(3)=6.403, p=.094$) and window 4 1001-1500ms ($\chi^2(3)=7.612, p=.055$) approached significance. The model for window 7 3001-3500ms was significant, $\chi^2(3)=9.930, p=.019$. The onset of the subordinate verb, the point of comparison in the present study, occurred on average in the fourth window 1501-2000ms ($M=1913.33, SD=457.83$). The significant window, window 7 3001-3500ms, occurred roughly 1000ms after the onset of the subordinate verb where increased processing difficulty with object-relatives is expected given the failure to associate the filler with the expected, non-present gap. This is consistent with Just and Carpenter’s (1993) finding that dilation maxima measured from the onset of the matrix verb (after the subordinate clause had ended) occurred on at average 1074ms for object-relatives and 958ms for subject-relatives.

<table>
<thead>
<tr>
<th>Window</th>
<th>Time (ms)</th>
<th>Chi-squared</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-500</td>
<td>3.142</td>
<td>.370</td>
</tr>
<tr>
<td>2</td>
<td>501-1000</td>
<td>3.934</td>
<td>.269</td>
</tr>
<tr>
<td>3</td>
<td>1001-1500</td>
<td>6.403</td>
<td>.094 †</td>
</tr>
<tr>
<td>4</td>
<td>1501-2000</td>
<td>7.612</td>
<td>.055 †</td>
</tr>
<tr>
<td>5</td>
<td>2001-2500</td>
<td>3.538</td>
<td>.316</td>
</tr>
<tr>
<td>6</td>
<td>2501-3000</td>
<td>5.030</td>
<td>.170</td>
</tr>
<tr>
<td>7</td>
<td>3001-3500</td>
<td>9.930</td>
<td>.019 *</td>
</tr>
<tr>
<td>8</td>
<td>3501-4000</td>
<td>4.900</td>
<td>.179</td>
</tr>
</tbody>
</table>

† $p<.1$  †† $p<.05$  ††† $p<.01$  †††† $p<.001$

Table 48: Significance of models by 500ms window. 3 degrees of freedom for all models.

The marginal means were calculated for each window and plotted with 95% confidence interval error bars in Figure 50. A horizontal, dashed green line is added at 6.25 diameter units to highlight the findings. For heritage speakers, dilation for subject-relatives and object-relatives are comparable at each window. Heritage speakers’ pupils dilate from the start of the trial to window 3 1001-1500ms and are stable until the end of the trial. Late bilinguals show differences between the subject-relatives and object-relatives. In response to object-relatives, late bilinguals’ pupils dilate from the start of the trial and peak in the window
In response to subject-relatives, late-bilinguals’ pupils dilate consistently from the start of the trial to the end of the trial albeit with a smaller slope than object-relatives.

### 6.4.1 Window 7: 3001-3500ms

Object-relatives elicited greater average dilation in the 3001-3500ms time window than subject-relatives for both groups. Mean window-average dilation by bilingual group and condition is plotted with 95% confidence interval error bars in Figure 51. Mean window-average dilation by bilingual group and condition is reported in Table 49.

Across group, average dilation in the 3001-3500ms time window for object-relatives ($M=7.14$, $SD=11.78$) was greater than subject-relatives ($M=6.23$, $SD=11.66$) indicating that across group, object-relatives were associated with greater cognitive effort 3 seconds after the onset of the trial. Late bilinguals appeared to experience slightly greater difficulty with all relative clause items. Late bilinguals ($M=8.05$, $SD=10.51$) had greater average dilation in the 3001-3500ms time window than heritage speakers ($M=6.14$, $SD=10.89$). This is possibly attributable to the large asymmetry between subject- and object-relatives. Only late bilinguals exhibited the subject-object processing asymmetry as indexed by average dilation in the 3001-3500ms time window. Late bilinguals had greater average dilation in the 3001-3500ms time window for object-relatives ($M=9.97$, $SD=9.78$) than subject-relatives ($M=6.14$, $SD=10.89$). Heritage speakers only had a very small difference in average dilation in the 3001-3500ms time window for object-relatives ($M=6.25$, $SD=12.43$) and subject-relatives ($M=6.27$, $SD=12.01$). Heritage speakers had slightly greater average dilation in the 3001-3500ms time window for subject-relatives than late bilinguals had for subject-relatives.

To determine whether differences between condition and bilingual group were statistically significant, average dilation in the 3001-3500ms time window was modeled. The maximal model is shown in Table 50 and includes the predictor variables of group, condition, and the interaction, controlling for working memory, had a significantly better fit than the working memory only model ($\chi^2(3)=9.930$, $p=.019$) and a simpler model without the interaction ($\chi^2(1)=6.317$, $p=.012$).
Figure 51: Mean average dilation in window 7 (3001-3500ms) by bilingual group and condition with 95% confidence interval error bars.

Table 49: Mean average dilation in window 7 (3001-3500ms) by bilingual group and condition. (Standard deviations in parentheses).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Late Bilinguals</th>
<th>Heritage Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-Relatives</td>
<td>6.14</td>
<td>6.27</td>
<td>6.23</td>
</tr>
<tr>
<td></td>
<td>(10.89)</td>
<td>(12.01)</td>
<td>(11.66)</td>
</tr>
<tr>
<td>Object-Relatives</td>
<td>9.97</td>
<td>6.25</td>
<td>7.41</td>
</tr>
<tr>
<td></td>
<td>(9.78)</td>
<td>(12.43)</td>
<td>(11.78)</td>
</tr>
<tr>
<td>Total</td>
<td>8.05</td>
<td>6.26</td>
<td>6.82</td>
</tr>
<tr>
<td></td>
<td>(10.51)</td>
<td>(12.21)</td>
<td>(11.73)</td>
</tr>
</tbody>
</table>

Table 50: Maximal linear mixed-effects model of average dilation in window 7 (3001-3500ms) elicited by subject- and object-relative clauses.

The intercept of the model was not significant, meaning that late bilingual average dilation in the 3001-3500ms time window for subject-relatives did not significantly differed from the mean across condition.
and across group, controlling for working memory ($M=6.82$, $SD=11.73$), $B=4.340$, $SE(B)=4.809$, $t(33.1)=0.903$, $p=.373$. The predictor variable of group was not significant, meaning that although heritage speakers had more average dilation in the 3001-3500ms time window for subject-relatives than late bilinguals did for subject-relatives, this difference was not significant, $B=0.325$, $SE(B)=1.843$, $t(46.6)=0.176$, $p=.861$. The predictor variable of condition was significant, meaning that late bilinguals had significantly more average dilation in the 3001-3500ms time window for object-relatives than subject-relatives, $B=3.823$, $SE(B)=1.266$, $t(928.0)=3.021$, $p=.003$. Lastly, the interaction between bilingual group and condition was significant meaning that the difference between subject-relatives and object-relatives was significantly different between bilingual group, $B=-3.843$, $SE(B)=1.526$, $t(928.0)=-2.518$, $p=.012$. The working memory control variables as measured by the n-back ($B=-0.983$, $SE(B)=1.164$, $t(32.0)=-0.844$, $p=.405$) and Spanish serial recall ($B=0.661$, $SE(B)=0.839$, $t(32.0)=0.788$, $p=.436$) tasks, were not significant predictors of window-average TEPR in the 3001-3500ms time window.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>CI_lower</th>
<th>CI_upper</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late Bilingual</td>
<td>7.917</td>
<td>1.348</td>
<td>32.00</td>
<td>5.097</td>
<td>10.737</td>
</tr>
<tr>
<td>Heritage Speaker</td>
<td>6.321</td>
<td>0.928</td>
<td>32.00</td>
<td>4.430</td>
<td>8.211</td>
</tr>
<tr>
<td><strong>Relative Clause</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>6.168</td>
<td>0.912</td>
<td>46.96</td>
<td>4.334</td>
<td>8.002</td>
</tr>
<tr>
<td>Object</td>
<td>8.070</td>
<td>0.912</td>
<td>46.96</td>
<td>6.236</td>
<td>9.904</td>
</tr>
<tr>
<td><strong>Late Bilingual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject-relatives</td>
<td>6.006</td>
<td>1.522</td>
<td>46.70</td>
<td>2.943</td>
<td>9.068</td>
</tr>
<tr>
<td>Object-relatives</td>
<td>9.829</td>
<td>1.522</td>
<td>46.70</td>
<td>6.766</td>
<td>12.891</td>
</tr>
<tr>
<td><strong>Heritage Speaker</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject-relatives</td>
<td>6.331</td>
<td>1.021</td>
<td>46.88</td>
<td>4.275</td>
<td>8.386</td>
</tr>
<tr>
<td>Object-relatives</td>
<td>6.311</td>
<td>1.021</td>
<td>46.88</td>
<td>4.256</td>
<td>8.366</td>
</tr>
</tbody>
</table>

Table 51: Predicted marginal means from maximal model of average dilation in window 7 (3001-3500ms) elicited by subject- and object-relative clauses.

The predicted values from the maximal model were used to calculate descriptive statistics, shown in Table 51 and as outlined in section 5.3, to compare groups and explore the subject-object asymmetry by-and across groups after having controlled for individual- and item-level variance and working memory. In an opposite trend than the other dependent variables, across condition, late bilinguals ($M_p=7.917$, $SE_p=1.348$, 95% $CI_p$ [5.097, 10.737]) had more average dilation in the 3001-3500ms time window than heritage speakers ($M_p=6.321$, $SE_p=0.928$, 95% $CI_p$ [4.430, 8.211]) although this difference was not significant. Across group, object-relatives ($M_p=8.070$, $SE_p=0.912$, 95% $CI_p$ [6.236, 9.904]) elicited
significantly more window-average dilation than subject-relatives ($M_p=6.168$, $SE_p=0.912$, 95% CI $[4.334, 8.002]$). For late bilinguals, object-relatives ($M_p=9.829$, $SE_p=1.522$, 95% CI $[6.766, 12.891]$) elicited significantly more average dilation in the 3001-3500ms time window than subject-relatives ($M_p=6.006$, $SE_p=1.522$, 95% CI $[2.943, 9.068]$). For heritage speaker participants, although subject-relatives ($M_p=6.331$, $SE_p=1.021$, 95% CI $[4.256, 8.366]$) elicited slightly more average dilation in the 3001-3500ms time window than object-relatives ($M_p=6.311$, $SE_p=1.021$, 95% CI $[4.275, 8.386]$), this difference was not significant. Heritage speakers did not have significantly more average dilation in the 3001-3500ms time window for subject-relatives that late bilinguals did for subject-relatives.

<table>
<thead>
<tr>
<th></th>
<th>r-value</th>
<th>p-value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maximum dilation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~ Trial Average</td>
<td>.385</td>
<td>.011</td>
<td>[.096, .614]</td>
</tr>
<tr>
<td>~ Epoch Average</td>
<td>.440</td>
<td>.003</td>
<td>[.161, .654]</td>
</tr>
<tr>
<td>~ 3001-3500ms</td>
<td>.349</td>
<td>.022</td>
<td>[.054, .588]</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial ~ Epoch</td>
<td>.865</td>
<td>&lt;.001</td>
<td>[.763, .925]</td>
</tr>
<tr>
<td><strong>3001-3500ms window</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~ Trial Average</td>
<td>.689</td>
<td>&lt;.001</td>
<td>[.491, .820]</td>
</tr>
<tr>
<td>~ Epoch Average</td>
<td>.879</td>
<td>&lt;.001</td>
<td>[.785, .933]</td>
</tr>
</tbody>
</table>

Table 52: Dependent variable correlations across bilingual group.

6.5 TEPR comparison

Across bilingual group, all of the TEPR difference amplitude variables were significantly correlated with each other, in Table 52. The average TEPR variables (trial-average, epoch-average, window-average) were all strongly correlated with each other and there was only a medium correlation between the average TEPR variables and maximum TEPR. Trial-average TEPR and epoch-average TEPR were significantly, strongly correlated with each other, $r(41)=.865$, 95% CI $[.763, .925]$, p<.001. Window-average TEPR from window 7 3001-3500ms was significantly correlated with epoch-average TEPR ($r(41)=.879$, 95% CI $[.785, .933]$, p<.001) and significantly less so with trial-average TEPR ($r(41)=.689$, 95% CI $[.491, .820]$, p<.001). This makes sense as values in window 7 3001-3500ms were included in the epoch-average TEPR.
Maximum dilation was significantly correlated with trial-average TEPR ($r(41)=.385$, 95% CI [.096, .614], $p=.011$), epoch-average TEPR ($r(41)=.440$, 95% CI [.161, .654], $p=.003$), and window 7 3001-3500ms average TEPR ($r(41)=.349$, 95% CI [.054, .588], $p=.022$), although to a significantly less degree than the correlations between the average TEPRs. These correlations suggest that maximum TEPR is distinct from the average TEPR variables.

For late bilinguals, the average TEPR variables were all strongly correlated with each other in Table 53 and illustrated in Figure 52. Trial-average TEPR and epoch-average TEPR were significantly, strongly correlated with each other, $r(15)=.861$, 95% CI [.649, .949], $p<.001$. Window-average TEPR from window 7 3001-3500ms was significantly correlated with epoch-average TEPR ($r(15)=.878$, 95% CI [.688, .956], $p<.001$) and less so with trial-average TEPR ($r(15)=.694$, 95% CI [.320, .881], $p=.002$). As with the across group correlations, this makes sense as values in window 7 3001-3500ms were included in the epoch-average TEPR.

There was a significant correlation between maximum TEPR and epoch-average TEPR ($r(15)=.489$, 95% CI [.011, .785], $p=.046$), and window 7 3001-3500ms average TEPR ($r(15)=.529$, 95% CI [.065, .805], $p=.029$), respectively. Maximum TEPR was not significantly correlated with trial-average TEPR ($r(15)=.248$, 95% CI [-.264, .651], $p=.338$), similar to Ledoux and colleagues (2016). However, the significant correlations between maximum TEPR and average TEPR suggests that the dilation maxima occurred in the epoch region around 3001-3500ms.

<table>
<thead>
<tr>
<th></th>
<th>r-value</th>
<th>p-value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum dilation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~ Trial Average</td>
<td>.248</td>
<td>.338</td>
<td>[.264, .651]</td>
</tr>
<tr>
<td>~ Epoch Average</td>
<td>.489*</td>
<td>.046</td>
<td>[.011, .785]</td>
</tr>
<tr>
<td>~ 3001-3500ms Window</td>
<td>.529*</td>
<td>.029</td>
<td>[.065, .805]</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial ~ Epoch</td>
<td>.861***</td>
<td>&lt;.001</td>
<td>[.649, .949]</td>
</tr>
<tr>
<td>3001-3500ms window</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~ Trial Average</td>
<td>.694**</td>
<td>.002</td>
<td>[.320, .881]</td>
</tr>
<tr>
<td>~ Epoch Average</td>
<td>.878***</td>
<td>&lt;.001</td>
<td>[.688, .956]</td>
</tr>
</tbody>
</table>

Table 53: Dependent variable correlations for late bilinguals.
Similar to late bilinguals, the average TEPR variables were all strongly correlated with each other for heritage speakers in Table 54 and illustrated in Figure 53. Trial-average TEPR and epoch-average TEPR were significantly, strongly correlated with each other, \( r(24)=.839, \) 95% CI [0.670, 0.926], \( p<.001 \). Window-average TEPR from window 7 3001-3500ms was significantly correlated with epoch-average TEPR \( (r(24)=.844, \) 95% CI [0.679, 0.928], \( p<.001 \)) and significantly less so with trial-average TEPR \( (r(24)=.605, \) 95% CI [0.284, 0.804], \( p=.001 \)). As with the across group correlations and with late bilingual correlations, this makes sense as values in window 7 3001-3500ms were included in the epoch-average TEPR.

<table>
<thead>
<tr>
<th>r-value</th>
<th>p-value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum dilation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sim ) Trial Average</td>
<td>.397*</td>
<td>.045 [.012 .680]</td>
</tr>
<tr>
<td>( \sim ) Epoch Average</td>
<td>.272</td>
<td>.180 [-.129 .596]</td>
</tr>
<tr>
<td>( \sim ) 3001-3500ms</td>
<td>.019</td>
<td>.926 [-.371 .404]</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial ( \sim ) Epoch</td>
<td>.839***</td>
<td>&lt;.001 [.670 .926]</td>
</tr>
<tr>
<td>3001-3500ms window</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sim ) Trial Average</td>
<td>.605**</td>
<td>.001 [.284 .804]</td>
</tr>
<tr>
<td>( \sim ) Epoch Average</td>
<td>.844***</td>
<td>&lt;.001 [.679 .928]</td>
</tr>
</tbody>
</table>

Table 54: Dependent variable correlations for heritage speakers.

In an opposite pattern from late bilinguals, heritage speaker maximum TEPR was significantly correlated with trial-average TEPR and not the later occurring average regions. There was a significant, medium
correlation between maximum TEPR and trial-average TEPR, \( r(24) = .397, 95\% CI [0.012, .680], p = .045 \). Maximum TEPR was not significantly correlated with epoch-average TEPR (\( r(24) = .272, 95\% CI [-.129, .596], p = .180 \)), and window 7 3001-3500ms average TEPR (\( r(24) = .019, 95\% CI [-.371, .404], p = .926 \)), respectively. These correlations between maximum TEPR and average TEPRs suggests that the dilation maxima occurred before the epoch region. It is possible that heritage speakers had increased processing load earlier than late bilinguals that was unrelated to the subject-object relative clause processing asymmetry.

7. Discussion

Previous pupillometry literature found that object-relatives elicited greater pupillary responses than subject-relatives. The findings in the present study are consistent with previous literature from monolinguals for late bilinguals only. The TEPR effects by bilingual group in the current study are in Table 55. Significant observed effects are indicated with a tick mark <✓>. For late bilinguals, object-relatives elicited significantly greater TEPR amplitude than subject-relatives. No significant differences between relative clause condition was observed for heritage speakers.

<table>
<thead>
<tr>
<th>Task-evoked Pupillary Response</th>
<th>Late Bilinguals</th>
<th>Heritage Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>Trial</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Epoch</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>3001-3500ms</td>
<td>✓</td>
</tr>
<tr>
<td>Maximum</td>
<td>✓</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 55: TEPR results

The first research question was: Are TEPR effects components that indicate processing difficulty greater for object-relatives than subject-relatives in late bilinguals? We can confirm our expectation that object-relatives are indexed by greater processing load than subject-relatives in late bilinguals. Our second research question was: Do heritage speakers process object-relatives and subject-relatives differently when compared to late bilinguals when hearing Spanish sentences? We cannot confirm our initial
hypothesis that heritage speakers will process relative clauses similar to late bilinguals. In all cases, heritage speakers differed from late bilinguals in processing relative clauses. Object-relatives did not elicit greater processing load than subject-relatives for heritage speakers. The findings from the correlations suggest that heritage speakers had greater TEPR amplitude earlier than late bilinguals and displayed greater processing load for both object-relatives and subject-relatives.

7.1 Lack of TEPR effect for heritage speakers

The effect observed for late bilinguals is consistent with previous literature. As anticipated, greater TEPR amplitude is found in response to the object-relative than subject-relative reflecting processing difficulty arising from syntactic reanalysis and/or maintaining the filler in working memory. No difference in TEPR amplitude difference between object- and subject-relatives was observed for heritage speakers. There are a number of possible reasons for the absence of a subject-object relative clause processing asymmetry in heritage speakers, such as abandonment of processing, disengagement from the task, group differences in working memory, and processing overload. These possible sources are dismissed, and we propose that the lack of an observed processing asymmetry results from greater processing efficiency in heritage speakers as a result of their unique bilingual experience.

Processing load has been found to decrease once individuals are no longer required to hold items in working memory (Kahneman & Beatty, 1966; Johnson, 1971; Fernandez, 2013). If heritage speakers abandoned processing of the relative clauses due to difficulty, thereby no longer holding the filler in working memory, this would have been evident as decreased pupil size. This was not observed. Heritage speakers did not demonstrate a decrease in pupil size for neither subject- nor object-relatives during the trials. Pupil size for both conditions reached asymptote and did not decrease.

It is also possible that the lack of increased processing load for the object-relatives resulted from task disengagement. Disengagement from task (Smallwood et al., 2012; Zekveld & Kramer, 2014) or mind-wandering (Gilzenrat et al., 2010) has been shown to lead to decreased TEPR responses to increasing task difficulty. The present study utilizes a similar protocol to ERP studies reported in chapter 3 and in Martohardjono and colleagues (2017a, 2017b) who found significant ERP effects of processing for other conditions in the project. ERP components also attenuate due to mental fatigue (Boksem, Meijman,
& Lorist, 2005; Ansorge et al., 2011) and task disengagement (Hopstaken et al., 2015) thereby making it unlikely that participants were disengaging during the present study just for the relative clause items.

Just & Carpenter (1993) found that mean pupillary responses elicited by relative clause items, including subject-relatives, for individuals with lower working memory capacity were greater, however, this difference was not significant. This is unlikely to be the case as there were no significant group differences in working memory (section 3.2.1) and working memory was included as a control variable in statistical models.

If heritage speakers have greater heritage language processing difficulty, it is likely that the increased processing load for object-relatives would not be evident if heritage speakers reached their processing capacity. While it has been reported that processing overload does not result in a decrease in pupil size (Peavler, 1974), more studies have found that a decrease in pupil size is observed once working memory capacity has been exceeded (Poock, 1973; Granholm et al., 1996; Gutiérrez & Shapiro, 2010 cited by Fernandez, 2013). It is more likely that the differences in processing between bilingual groups are related to processing efficiency than processing overload for heritage speakers.

Pupillometry is not an absolute measure of processing demands on cognition but instead has been argued to index efficiency (Just, Carpenter, & Miyake, 2003). In mathematical tasks, individuals with higher intelligence exhibited smaller pupil dilations. This is argued to reflect more efficient processing, not less effort or mental energy to process information (Ahern & Beatty, 1979). Higher intelligence individuals are better able to modulate resource allocation given task demands (Lee et al., 2015) and mobilize more processing resources for harder tasks (van der Meer et al., 2010; Wendt, Dau, & Hjortkjær, 2016). Additionally, learning has been shown to attenuate pupil size regardless of task difficulty (Sibley, Coyne and Baldwin 2011) and galvanic skin response (Brown, 1937; Kintsch, 1965), a related autonomic response indexing locus coeruleus-norepinephrine system activation (Tyler et al., 2015). Further practice (use of language two languages) could result in more automatic and efficient processing thereby decreasing processing load as indexed by autonomic responses.

We are not arguing that heritage speakers are more intelligent or have learned Spanish better than late bilinguals, although we did not measure intelligence, but instead are hypothesizing that heritage speakers are likely more efficient in their processing. Piquado, Isaacowitz, and Wingfield (2010) found
relative clause processing differences between older and younger adults, which they argued to be an
effect of efficiency. They found that older adults did not evidence the processing asymmetry in less
complex sentences and only in more complex sentences, reflecting more efficiency in processing. In older
adults, they argue that “the cognitive effort required for dealing with online interpretive processing does
not exceed a threshold necessary to yield a measurable change in pupil size” (Piquado, Isaacowitz, &
Wingfield, 2010, p. 12). Increased efficiency for heritage speakers could result in a lack of observed
subject-object relative clause processing asymmetry. Sources of more efficient processing could be
related to the experience of growing up in a bilingual environment; heritage speakers have lived their
whole life in bilingual context. The additional processing demands involved in bilingualism could result in
more efficient language processing.

7.2 Heritage speaker language experience

Although heritage speakers are native, L1 speakers of Spanish, the observed differences in processing
(greater processing difficulty for subject-relatives or more efficient processing strategies for both subject-
and object-relatives) could be attributed to differences in language use and exposure. Greater processing
load has been observed in bilinguals when processing their non-native (Hyönä, Tommola, & Alaja, 1995;
Borghini & Hazan, 2018) or non-dominant language (Byers-Heinlein, Morin-Lessard, & Lew-Williams,
2017). Schmidtke (2014) found that lower proficiency bilinguals had greater pupillary responses which
they argued was the effect of language experience. However, Schmidtke (2014) also observed that early
bilinguals had similar to pupillary responses to monolinguals when processing of L2. While proficiency
could be driving the observed effect in the present study in subject-relative processing in heritage
speakers, it is unlikely. Heritage speakers did not score differently than late bilinguals on the Spanish
RMST (Klein & Martohardjono, 2009) as reported in section 3.2.2 which included relative clause
structures indicating competence in Spanish syntax. Likewise, heritage speakers did not differ from late
bilinguals in oral Spanish fluency as reported in section 3.2.3. However, heritage speakers in the present
study, and in previous studies, vary from late bilinguals in terms of language use (use Spanish in different
domains), and language exposure (Spanish is acquired under different contexts). Heritage speakers also
differ in self rated language ability, however, this measure does not reliably index language proficiency.
Differences between English language exposure, current Spanish language use, and self-reported language abilities are explored further.

### Table 56: Mean percent of English use for language exposure variables (Standard deviations in parentheses; Significance of group differences from t-test indicated)

<table>
<thead>
<tr>
<th>English Exposure</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Late Bilinguals (%)</td>
</tr>
<tr>
<td></td>
<td>(Standard deviations in parentheses)</td>
</tr>
<tr>
<td>Cumulative</td>
<td>12.45 (12.00)</td>
</tr>
<tr>
<td>English as LOLT</td>
<td>12.82 (25.85)</td>
</tr>
<tr>
<td>English in Community</td>
<td>1.68 (3.12)</td>
</tr>
</tbody>
</table>

Table 56: Mean percent of English use for language exposure variables (Standard deviations in parentheses; Significance of group differences from t-test indicated)

<table>
<thead>
<tr>
<th>Current Spanish Use</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Late Bilinguals (%)</td>
</tr>
<tr>
<td></td>
<td>(Standard deviations in parentheses)</td>
</tr>
<tr>
<td>Domains</td>
<td>38.54 (12.12)</td>
</tr>
<tr>
<td>With Family</td>
<td>81.56 (14.66)</td>
</tr>
<tr>
<td>With non-Family</td>
<td>32.55 (19.56)</td>
</tr>
<tr>
<td>Media</td>
<td>30.73 (17.93)</td>
</tr>
</tbody>
</table>

Table 57: Mean percent of Spanish use for current language use variables (Standard deviations in parentheses; Significance of group differences from t-test indicated)

Three language exposure variables looked at self-reported exposure to English cumulatively over the participants' lifetime, English-medium education, and English used in the community of the participant from ages 5-18. Heritage speakers reported significantly more exposure to English than late bilinguals, as can be seen in Table 56. Cumulative English exposure is operationalized as a percentage of participant's life spend in an English-dominant environment (either starting a school where English is the language of
learning and teaching or age of arrival to the mainland anglophone US). In terms of a percentage of the participant’s life, heritage speakers ($M=84.47\%, SD=9.46$) reported being in an English-dominant environment for significantly larger percentage of their life than late bilinguals ($M=12.45\%, SD=12.00$), $t(41)=-21.95, p<.001, r=.96$. A significantly larger percentage of heritage speakers’ education ($M=93.96\%, SD=12.85$) was in English-medium classrooms than late bilinguals ($M=12.82\%, SD=25.85$), $t(21.23)=-12.01, p<.001, r=.93$. Similarly, a significantly larger percentage of heritage speakers’ youth ($M=63.60\%, SD=41.75$) between the ages of 5 and 18 were spend in English-dominant communities than late bilinguals ($M=1.68\%, SD=3.12$), $t(25.43)=-7.53, p<.001, r=.83$.

There were also group differences in the percentage of Spanish participants reported using currently although in certain domains both groups used equivalent amounts of Spanish, as can be seen in Table 57. Heritage speakers ($M=29.81\%, SD=12.06$) reported using significantly less Spanish in the home, social setting, and at work than late bilinguals ($M=38.54\%, SD=12.12$), $t(40)=2.27, p=.03, r=.34$. Heritage speakers ($M=48.01\%, SD=17.24$) reported using significantly Spanish with family members (i.e., mother, father, siblings, younger children, partner) than late bilinguals ($M=81.56\%, SD=14.66$), $t(40)=6.47, p<.001, r=.72$. This trend also extends to non-family members (i.e., friends, boss, co-workers, classmates). Heritage speakers ($M=18.11\%, SD=15.72$) reported using significantly less Spanish with non-family members than late bilinguals ($M=32.55\%, SD=19.56$), $t(40)=2.63, p=.01, r=.38$. Both groups reported using Spanish when listening to music, reading, and watching TV equivalently. The percentage of the time that US Latinx bilinguals read in Spanish, watched Spanish-language TV, and listened to Spanish-language music did not significantly differ (Heritage speakers: $M=28.53\%, SD=13.58$; Late bilinguals: $M=30.73\%, SD=17.93$), $t(40)=0.45, p=.65, r=.07$.

The groups also differed in terms of self-reported Spanish and English language ability in Table 58, although this is not surprising given the well-documented effects of linguistic insecurity in heritage speakers. Participants were asked to rate their language ability on a 5-point Likert scale with the endpoints labeled 1=limited knowledge to 5=native. Late bilinguals reported higher Spanish proficiency in both literacy and speaking. All late bilinguals reported their Spanish literacy skills as native. Heritage speakers, however, reported the Spanish literacy skills ($M=3.73, SD=0.92$) as significantly different than native, $t(25)=-7.04, p<.001, r=.82$. Late bilinguals ($M=4.97, SD=0.12$) self-reported significantly higher
Spanish speaking skills than heritage speakers ($M=4.50$, $SD=0.57$), $t(28.42)=4.10$, $p<.001$, $r=.61$.

Heritage speakers reported higher English proficiency in both literacy and speaking. All heritage speakers reported their English speaking and literacy skills as native. Late bilinguals, however, reported their English literacy skills ($M=4.06$, $SD=0.68$) and English speaking skills ($M=4.03$, $SD=0.60$) as significantly different than native, English Literacy: $t(15)=-5.51$, $p<.001$, $r=.82$; English Speaking: $t(16)=-6.68$, $p<.001$, $r=.86$.

<table>
<thead>
<tr>
<th>Language Ability</th>
<th>Group</th>
<th>Late Bilinguals (%)</th>
<th>Heritage Speakers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Spanish Literacy</td>
<td></td>
<td>5.00</td>
<td>(0.00)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.97</td>
<td>(0.12)</td>
</tr>
<tr>
<td>English Literacy</td>
<td></td>
<td>4.06</td>
<td>(0.68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.03</td>
<td>(0.60)</td>
</tr>
</tbody>
</table>

Table 58: Mean percent of language ability by language by skill (Standard deviations in parentheses; Significance of group differences from t-test indicated; 1=limited knowledge, 5=native)

Heritage speakers in the present study have spent more of their life in an English-dominant environment, were educated in English and use more English when communicating with family and non-family alike than late bilinguals. This dominance in English likely impacts processing of their native language. Further analyses are needed to explore the exact effect of language use and exposure in Spanish-English bilinguals to determine if group differences are attributable to variations in how the two groups use language. Likewise, differences in proficiency ought to be further explored, however, a measure of proficiency other than self-rating are needed due to linguistic insecurity.
7.3 Future research

The linguistic insecurity of heritage speakers could possibly result in anxiety during experimental investigations of their L1. Increased pupil size has been observed in response to anxiety (Johnson, 1971) and positive/negative emotions (van Steenbergen, Band, & Hommel, 2011). The emotional state or anxiety about heritage speakers’ non-dominant language was not measured in the present study. Future study is needed to operationalize and explore the effect of linguistic anxiety in language processing.

In the present study, difficulty resulting from syntactic reanalysis or working memory demands are similar to previous studies using subject-embedded relative clauses, however, difficulty arising from structural or thematic mismatch are different for object-embedded relative clauses. The head noun in the current study is the matrix object which would result in the subject-relative being harder to process due to the mismatch between the head nouns thematic or structural role in the matrix clause and subordinate clause (Traxler, Morris, & Seely, 2002). While no present studies to have explored the effect of thematic mismatch using pupillometry, a P300 ERP effect has been found (Bornkessel, et al., 2003). The P300 has been linked to locus coeruleus-norepinephrine system activation (Nieuwenhuis, de Geus, & Aston-Jones, 2013) which can be measured as increased pupil size. If thematic mismatch resulted in increased dilation, this could contribute to a lack of effect for heritage speakers if they are more sensitive to thematic mismatch. If there was greater dilation for subject-relatives resulting from thematic mismatch and greater dilation for object-relatives resulting from syntactic reanalysis or working memory demands, then no asymmetry would be observed. To explore the effect of thematic mismatch in heritage speakers versus late bilinguals, a follow-up study is needed which uses subject-embedded subject- and object-relatives.

8. Summary

Heritage speakers process relative clauses differently than late bilinguals in their first-learnt language. The present study avoids the confounds of linguistic insecurity and cross-linguistic structural effects, however, differences between heritage speakers and late bilinguals were observed in native language processing which likely result from differences in processing efficiency. Greater TEPR amplitude was observed for object-relatives than subject-relatives for late bilinguals, consistent with previous studies of
monolinguals. The processing differences observed are likely rooted in different patterns of language use and exposure to the HL and ML for heritage speakers.
CHAPTER VI:
CONCLUSION

In this chapter, the findings of the present study are presented and discussed. Section one restates the results of the five studies presented in previous chapters. Section two reviews the main research questions of the dissertation and discusses the findings in relation to the primary research questions. Section three compares the results of the event-related potentials and task-evoked pupillary response experiments with a combinatoric study of the probability of observing the effects in a small number of shared data in late bilinguals. Section four explores the possible sources of heritage speaker processing differences. A model of heritage speaker processing that accounts for the additional cognitive processes concurrent with relative clause processing is proposed in section five. Section six makes concrete recommendations for the effective study of heritage speakers. Lastly, section seven proposes future studies to clarify and further explore the findings of the present study.

1. Results
The findings from the 5 studies conducted in previous chapters are summarized below.

1.1 Spanish and English exposure, use, ability and identity
The Spanish and English language exposure, use, ability, and identity of US Latinx bilinguals in New York City were explored. Survey responses from a language background questionnaire administered to second-generation Latinx Americans (bilinguals who were either born in the anglophone US or were brought to the anglophone US before age 9) and first-generation Latinx Americans (bilinguals who moved to the anglophone US after age 18). The second-generation participants were heritage speakers of Spanish who were raised by Spanish-speaking, first-generation Latinx Americans. The first-generation participants spoke Spanish as a first language and were not immersed in an English dominant society until adulthood. Their survey responses were combined, and a principal component analysis was run to identify the independent constructs/behaviors present in the data. Components were: 1) language of work/public spaces; 2) language of media; 3) general English use/exposure; 4) English native speaker
identity; 5) Spanish native speaker identity; 6) English speaking and literacy ability; and 7) Spanish speaking ability.

Component scores between group were tested to identify areas of similarities and differences between the heritage speakers and late bilinguals. Heritage speakers and late bilinguals were similar in language use, exposure, and identity for some of these distinct constructs, and were similar in language ability and language use associated with ability. Heritage speakers, who were raised in the anglophone US, used more English which speaking with non-family members (e.g., coworkers, classmates) and when watching TV or listening to the radio than late bilinguals did. Each group also differed in their identification with their dominant language: English for heritage speakers, Spanish for late bilinguals. Identity was also associated with the language Spanish-English bilinguals used with their family. Identification as a Spanish native speaker was associated with using more Spanish with one’s parents and identification as an English native speaker was associated with using more English with one’s siblings. Both types of bilinguals are fluent users of English and Spanish. They were similar in their self-rated abilities in English (speaking and literacy) and Spanish (speaking only) and the language they used with their friends and partner. It is not surprising that Spanish literacy ability did not pattern with Spanish speaking ability as heritage speakers were educated in English-medium schools, or that higher self-rated Spanish speaking ability was associated with more use of Spanish with friends and partners. This analysis demonstrates that while heritage speakers differed from their apparent-time parents in self-identification and language use, they did not differ in their command of both languages.

1.2 Acceptability judgement task

Performance on metalinguistic tasks were modeled using appropriate statistical techniques to explore the established “yes-bias”, “maybe-bias”, and hesitancy of heritage speakers when responding to grammaticality judgment tasks. Heritage speakers are compared to the appropriate baseline of late bilinguals. The zero-and-one inflated beta mixed-effects model analysis of naturalness ratings of syntactically complex Spanish sentences showed that heritage speakers do indeed rate items differently than late bilinguals, which can be argued to reflect linguistic insecurity when completing a task requiring metalinguistic judgments. Across nine different conditions, heritages speakers were less likely to find a
sentence fully natural and used the mid-point more than late bilinguals. In the present study, the midpoint of the Likert scale was labelled *es posible* ‘maybe’. Consequently, the linguistic insecurity of heritage speakers manifests as failure to fully accept or fully reject most items, and a bias towards rejecting items (including ungrammatical items). This “maybe-bias” is possibly either a function of heritage speakers’ linguistic insecurity or common edge-avoidance also referred to as “fence-sitting”. Previous researchers have conjectured that that insecurity leads heritage speakers to over-accept items as a “yes-bias”, however, the current findings suggest that the linguistic insecurity leads heritage speakers to be ambivalent about whether a sentence is ‘good’ but can still identify ‘bad’ sentences, in contrast to the predictions of a heritage speaker “yes-bias”.

Late bilinguals, however, did not respond as one might expect given the prescriptive grammaticality of the sentences. Overall, late bilinguals tended to be more accepting of items than heritage speakers. Late bilinguals were more likely to rate an item as fully natural, and less likely to rate an item as fully unnatural. The pattern of late bilinguals being more accepting of ungrammatical items and the pattern of heritage speakers being more likely to fully reject items is consistent with findings that individuals with high linguistic insecurity are more sensitive to norms (Labov, 1990).

The findings of the present study confirm the claim that metalinguistic tasks are inappropriate for heritage speakers, but also suggest that these tasks are inappropriate for late bilinguals as well. An additional interesting finding is that there was more consensus in rating grammatical items than ungrammatical items. This suggests that, while notions of naturalness or grammaticality converge, the relationship between naturalness and ungrammaticality is less clear. This may reflect the absence of an underlying cognitive representation for ungrammatical sentences (Sprouse, 2007b). Thus, naturalness ratings of ungrammatical items have no comparison point.

1.3 Relative clause comprehension

To ensure that the present study focus on relative clause processing was not confounded by a lack of sufficient control of relative clause structures in heritage speakers, an analysis of a picture-selection comprehension task was run. Both late bilinguals and heritage speakers were accurate in selecting the correct picture. There were no differences in accuracy between the bilingual group or for the two different
relative clause type structures. Consequently, the observed subject-object comprehension asymmetry in previous studies of adult heritage speakers (Polinsky, 2008, 2011; Lee-Ellis, 2011; Lee, 2013) is not supported.

1.4 Event-related potentials
To appropriately investigate language processing in heritage speakers, physiological responses to neural language processing are first measured as changes in the electrical field at the scalp. Previous research argued that the subject-object relative clause processing asymmetry was observable as LAN, N400, and/or P600. The present study made use of an under-utilized procedure, a temporo-spatial principal component analysis, that better models the time-course and spatial relations of neuroelectrical behavior in the identification of ERP components. No ERP components consistent with a LAN were identified. Two ERP components consistent with a P600 were identified: an anterior/frontal P600 and the classic centroparietal P600. Heritage speakers had significantly greater frontal P600 amplitude across condition than late bilinguals indicating greater processing difficulty on both object- and subject-relatives. Late bilinguals has significantly greater centroparietal P600 amplitude for object-relatives consistent with increased working memory and/or syntactic reanalysis of object-relatives. While two significant early, fast ERP components were found in the parietal region, the polarity of the effects differed between bilingual group. Heritage speakers had significant negative-going N400-like effect for object-relatives, and late bilinguals had positive-going P300-like effects. In both groups these fast, early effects were strongly significantly positively correlated with the centroparietal P600 effect. Late bilingual ERP results provide evidence of the subject-object relative clause processing asymmetry in L1-dominant bilinguals immersed in an L2-society. The P300/P400 effects are consistent with previous literature finding early, fast, positive-going deflections to violations of sequential expectancy (Dien et al., 2010), and reorganization/revision of one’s mental model (Duncan-Johnson & Donchin, 1977; Donchin, 1981; Aleksandrov & Maksimova, 1985; Donchin & Coles, 1988), which is required when processing object-relatives. Heritage speaker ERP results are argued to be evidence of processing difficulty. Greater centroparietal P600 amplitude on subject-relatives due to processing difficulty result in the lack of a centroparietal P600 effect for object-relatives. Object-relatives in heritage speakers elicit N400-like effects due to difficulty in syntactic
integration (King and Kutas, 1995) and are then attenuated as indicated by the later centroparietal P600 effect as expected (Brouwer et al., 2016; Brouwer & Crocker, 2017). Additionally, the attenuation of P300/P400 effects in heritage speakers is argued to reflect the increased working memory demands of the task for heritage speakers (Polich, 2007; Evans & Pollack, 2011).

1.5 Task-evoked pupillary responses

Again, to appropriately investigate language processing in heritage speakers, the physiological responses to neural language processing were also measured as changes in pupil diameter as a result of autonomic nervous system stimulation from the brain. Four dependent variables operationalizing task-evoked pupillary responses (TEPR) were calculated: trial-average amplitude, epoch-average amplitude, maximum (100ms window-average maximum) amplitude, and 500ms window-average amplitude. For the 500ms window-average amplitude, a window-by-window analysis was also run. Significant TEPR effects, trial-average amplitude, epoch-average amplitude, maximum amplitude, 3001-3500ms window-average amplitude, which index the subject-object relative clause processing asymmetry, were only found for late bilinguals. No subject-object relative clause processing asymmetry was observed for heritage speakers. Correlations between the TERP effects revealed that heritage speakers had earlier maximum effects than late bilinguals, whose maximum cognitive load was observed approximately 1000-1500ms after the critical gap (subject-relatives)/filled gap (object-relatives). Heritage speakers’ subject-relative TEPR amplitude was higher than subject-relative amplitude for late bilinguals suggesting the lack of an effect is a result of processing difficulty in heritage speakers.

2. Research questions

The two research questions of the dissertation are:

1. Do late bilinguals show psychophysiological responses related to greater processing difficulty for object-relatives than subject-relatives in their first-learnt language in:
   a. an event-related potentials (ERP) experiment, and
   b. a pupillometry experiment?
2. Do heritage speakers process object-relatives and subject-relatives differently when compared to late bilinguals in their first-learnt language in:
   a. an event-related potentials experiment, and
   b. a pupillometry experiment?

Utilizing an appropriate comparison group (late bilinguals) and investigating the processing of an appropriate linguistic phenomenon (subject-object relative clause processing asymmetry) answers our first research question in the affirmative. Late bilinguals, as expected, do indeed show psychophysiological responses related to greater processing difficulty for object-relatives than subject-relatives in their first-learnt language in an event-related potentials (ERP) experiment and a pupillometry experiment. The relationship between these psychophysiological measures in explored below.

As for our second research question, whether heritage speaker bilinguals process relative clauses in their heritage language similarly to late bilinguals, we did not find this to be the case. In all eight of the derived measures of ERP and TEPR responses, heritage speakers performed differently. The results of the ERP and TEPR studies converge. Heritage speakers do not demonstrate the well-established processing asymmetry for structures that they fully comprehend, which we argue reflects differences in processing efficiency. In both cases, subject-relatives were associated with greater amplitude resulting in the lack of an asymmetry. We argue that this reflects processing efficiency resulting from language dominance.

3. Psychophysiological measures

The findings of the implicit psychophysiological measures utilized in this study converge in both bilingual groups, however, only in late bilinguals were the neuroelectrical and pupillary responses significant. The four ERP effects (centroparietal P600, frontal P600, P375, P420) and four TEPR responses (trial-average, epoch-average, maximum, 3001-3500ms window-average) are directly compared across experiment. Unfortunately, there are data from only two late bilingual participants who completed both studies. Therefore, a combinatoric study is run that computes all possible combinations of ERP to TEPR responses to empirically establish the probability of observing the patterning in the results from the two
participants given the roughly 22 thousand other combinations that could have been observed. However, since this analysis is based on data from only 2 participants, the findings must be taken as only suggestive.

3.1 Empirical comparison of ERP and TEPR

Between the two studies, seventy-two participants were included in the present analysis. The ERP study was run first in 2013-2015, and despite attempts to recruit the same participants in the later pupillometry study in 2016-2018, only a small percentage (6.94%, n=5) returned to complete the second task. This, unfortunately, makes it difficult to statistically compare the results across measure to explore the relationship between pupillometry and ERP results. As presented above, we observed a significant subject-object relative clause processing asymmetry in late bilinguals for three of the ERP components (i.e., centroparietal P600, P375, P420) and all four of the TEPR dependent variables (i.e., trial-average, epoch-average, maximum, 3001-3500ms window-average). Two late bilingual participants, illustrated in Figure 54, completed both tasks and their data can be used to calculate the probability of observing ERP-pupillometry effects for these two participants by calculating all possible combinations for ERP-pupillometry effects for any two randomly selected participants across bilingual group. These possible combinations of participants ERP-pupillometry effects yields an empirical distribution that can be used to calculate the two-tailed probability of observing the actual combination of the observed two participants from all possible combinations. As illustrated in Figure 55, if we have ERP data for four participants (i.e., A, B, C, D) and we have pupillometry data from four participants (i.e., A, B, 1, 2) with participants A and B participating in both experiments, we have the observed dyad of ERP-pupillometry data from participants A and B. If we compute all the combination of participants from both studies, we have 15 data pairings. We can then combine these ERP-pupillometry pairs into 240 pair-dyads, excluding dyads where the pairs are the same. The 240 pair-dyads are then used to calculate an empirical distribution which is used to derive the two-tailed p-value of observing the pair-dyad AA-BB given the other random combinations that might have been possible.
Figure 54: Pupillometry and ERP data from late bilingual participants.

Figure 55: Illustration of possible combinations of participant data.
In the present study, we calculate the probability of observing the ERP-pupillometry effects found in two late bilingual participants out of the possible 20+ thousand non-repeating random pair-dyads \((N=22,050)\). This was not done for the three heritage speaker participants since there was no observed significant ERP-pupillometry effects in heritage speakers. The ERP and pupillometry effects in the present study were z-score normalized across group participant-average difference amplitude. The difference between z-scored ERP difference amplitude and z-scored pupillometry difference amplitude was calculated for each ERP-pupillometry pair, then summed for the pair-dyads. The summed z-scored ERP-pupillometry difference was calculated for each of the 2.1 million pair dyads by each of the sixteen ERP-pupillometry effect pairs. A histogram of possible pair-dyad summed z-scored difference for the centroparietal P600 ERP effect and the 3000-3500ms window-average pupillometry effect in Figure 56 shows that the observed centroparietal P600-window-average TEPR effect is in the left-tail of the empirical distribution. The cumulative distribution frequency in Figure 57 shows the same effect.

**Figure 56**: Histogram of combinations of dyad summed difference between z-scored centroparietal P600 difference amplitude and z-scored TEPR difference amplitude for the 3000-3500ms time window. Observed summed difference for late bilingual participants indicated w
Figure 57: Empirical cumulative distribution frequency of combinations of dyad summed difference between z-scored centroparietal P600 difference amplitude and z-scored TEPR difference amplitude for the 3000-3500ms time window. Observed summed difference for late

<table>
<thead>
<tr>
<th>ERP</th>
<th>TEPR</th>
<th>p-value</th>
<th>log(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centroparietal P600</td>
<td>Trial Average</td>
<td>0.21</td>
<td>-0.68</td>
</tr>
<tr>
<td></td>
<td>Epoch Average</td>
<td>0.26</td>
<td>-0.59</td>
</tr>
<tr>
<td></td>
<td>3001-3500ms</td>
<td>0.14</td>
<td>-0.85</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>0.28</td>
<td>-0.55</td>
</tr>
<tr>
<td>P375</td>
<td>Trial Average</td>
<td>0.21</td>
<td>-0.68</td>
</tr>
<tr>
<td></td>
<td>Epoch Average</td>
<td>0.25</td>
<td>-0.60</td>
</tr>
<tr>
<td></td>
<td>3001-3500ms</td>
<td>0.14</td>
<td>-0.85</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>0.28</td>
<td>-0.55</td>
</tr>
<tr>
<td>P420</td>
<td>Trial Average</td>
<td>0.35</td>
<td>-0.46</td>
</tr>
<tr>
<td></td>
<td>Epoch Average</td>
<td>0.25</td>
<td>-0.60</td>
</tr>
<tr>
<td></td>
<td>3001-3500ms</td>
<td>0.14</td>
<td>-0.85</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>0.28</td>
<td>-0.55</td>
</tr>
</tbody>
</table>

Table 59: Two-tailed empirical p-value and log-transformed p-value of the observed summed difference for late bilingual participants from empirical frequency distribution for significant effects.

3.2 Results

The two-tailed probability was calculated using the stats::pnorm() function in R from the empirical frequency distribution for each of the sixteen ERP-pupillometry effect pairs. The two-tailed p-values in Tables Table 59-Table 60 are also presented as log transformed values. The log transformation normalizes the p-value scale to represent the way it is used in research, where differences between p=.1
and \( p = .99 \) are less important than differences between \( p = .1 \) and \( p = .05 \). The log transformed standard \( p \)-value benchmarks are \( \log(1) = 0.00, \log(.10) = -1.00, \log(.05) = -1.30, \log(.01) = -2.00, \log(.001) = -3.00 \). For the three significant ERP effects, in Table 59, none of the ERP-pupillometry effect pairs are significant using the .05 alpha-level, however, all \( p \)-values are lower \( (p \ range: [.14, .35]; \log(p) \ range: [-0.85, -0.46]) \) than those observed between the non-significant ERP effect and significant pupillometry effect \( (p \ range: [.40, .59]; \log(p) \ range: [-0.40, -0.23]) \), in Table 60.

<table>
<thead>
<tr>
<th>ERP</th>
<th>TEPR</th>
<th>( p )-value</th>
<th>( \log(p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal P600</td>
<td>Trial Average</td>
<td>0.50</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>Epoch Average</td>
<td>0.56</td>
<td>-0.25</td>
</tr>
<tr>
<td></td>
<td>3001-3500ms</td>
<td>0.40</td>
<td>-0.40</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>0.59</td>
<td>-0.23</td>
</tr>
</tbody>
</table>

*Table 60: Two-tailed empirical \( p \)-value and log-transformed \( p \)-value of the observed summed difference for late bilingual participants from empirical frequency distribution for significant effects.*

In examining the significant ERP effects, the early and late positive deflections are most related to the window-average TEPR effects \( (\text{average } p = .14; \log(p) = -0.85) \) and least related to the maximum TEPR effect \( (\text{average } p = .28, \log(p) = -0.55) \). The probability of observing a P600, P375, or P420 effect and a TEPR effect was lowest for the 3000-3500ms time window average. One would observe an effect in both ERP and pupillometry when there was not one fewer than 3 in 20 times. The significant pupillometry effects were least related to the non-significant frontal P600 effect \( (\text{all } p > .40, \log(p) > -0.40) \). This is not surprising given that there were no frontal P600 effects in late bilinguals. If lower or overlapping \( p \)-values had been found it would have presented a problem given the non-significant frontal P600 effect and eroded confidence in the pair-dyad combinatoric findings previously mentioned.

3.3 Discussion

While the findings of the combinatoric study were not significant, they are informative; however, they must be taken with the largest grain of salt as the analysis was only based on data from 2 participants. The significant TEPR effects were more related \( (\text{more likely to be observed using two-tailed probability}) \) to the significant ERP effects. This suggests that, as expected, the additional cognitive processes involved in
processing object-relatives compared to subject-relatives as indexed by ERP components of syntactic reanalysis are most closely related to later TEPR measures and not measures of maximum cognitive effort. To test these predictions, an additional study is needed recording EEG and pupil diameter measures from the same participant simultaneously.

4. Factors impacting heritage speaker processing

Evidence from the present studies of relative clauses provide strong evidence that language processing is different in heritage speakers. The results of the ERP study implicate processing difficulty – increased processing costs for subject-relatives – as the reason for the lack of asymmetry observed in subject-object relative clause processing. The pupillometry results are similar in the lack of an observed subject-object relative clause processing asymmetry; however, taken together, the results suggest that additional factors are driving the lack of an effect. Indeed, heritage speakers are evidencing increases in processing costs. We argue that these are not due to difficulty in processing relative clauses, but rather to increased processing costs due to language dominance. We discuss possible factors driving the increased processing costs due to language dominance seen in heritage speakers, namely incomplete grammars, language activation, and heritage language anxiety.

4.1 Attrition and incomplete acquisition

The default view of heritage speaker studies takes the deficit perspective that heritage language grammars are incomplete, attrited, or simple. We argue that the findings of the present study, which possibly implicate processing difficulty as the source of our failure to observe the well-established and essentially cognitively primitive subject-object processing asymmetry, are not a result of attrition or incomplete acquisition.

Otheguy’s (2013) argument against incomplete acquisition, asks the reader to consider the definition of complete acquisition and what that would look like in the case of intergenerational change occurring in an undervalued variety. It further makes the point that the term heritage speakers of Spanish in the anglophone US certainly includes those speakers who successfully communicate in Spanish in many environments, such as talking to friends, family, and a diverse array of community interactions, but
can also include individuals who grow up with Latinx cultural practices without exposure to Spanish. The experiments in this investigation have selected heritage speakers based on the criteria of fluency in Spanish and English so that heritage speaker as a term is limited to only individuals who are fluent speakers of Spanish.

In order to have a valid understanding of the language competence and performance of heritage speakers, several points must be taken into consideration. These include the amount of Spanish and type of exposure to Spanish that heritage speakers throughout their lives, as well as the frequency of Spanish use and how that changes dramatically around age five. Putnam and Sanchez (2013) explore the claim of incomplete acquisition set forth by Montrul (2002, 2008, 2009) and Polinsky (1997, 2006). Without rejecting the claim of incomplete acquisition, Putnam and Sanchez (2013) raise the question of how much exposure to a linguistic feature is sufficient to be considered enough input to acquire complete acquisition of such a feature. In order to test the claim of incomplete acquisition in a bilingual sample, the point of comparison would necessarily be a group of bilinguals who are considered to have completely acquired both of their languages. Furthermore, the ultimate acquisition of both the L1 and L2 of said bilinguals would have to be consistent within this population in order for the notion of incomplete acquisition to be falsifiable. Although balanced bilinguals who have reached “native” levels of attainment in both of their languages likely exist, these do not represent the majority of the bilingual population.

According to Montrul (2002, 2008, 2009) and Polinsky (2006), critical periods of language acquisition mark the distinction between HL attrition and incomplete acquisition, in that HL attrition takes as its starting point a language system that has reached adult levels of nuance and grammar, but is then lost due to increased use of the ML; while the argument for incomplete acquisition is that due to the decreased use of the HL before the complete acquisition is achieved, there are grammatical forms and nuances that never fully develop. Most accepted theories of bilingual processing hypothesize that rather than imagining each of a bilingual’s grammars as a separate entity, the bilingual grammars have a bidirectional influence on each other and are managed by periods of activation and suppression. According to the model outlined by Putnam and Sánchez (2013), the difference between HL attrition and incomplete acquisition is therefore superficial. Heritage language speakers map formal linguistic features together in
ways that are not necessarily expected or predicted in monolingual variants of the HL but are expected and predicted by models that rely on frequency of language use.

One might argue that heritage speakers do not control relative clauses (in spite of the evidence of relative clause comprehension). Some researchers argue that the earlier a heritage speaker is exposed to the ML, the more likely HL skills will be weaker during the teenage years (Oh & Au, 2005; Montrul, 2008). However, this finding was not replicated in Kupisch and colleagues (2014), and the field is in need of a more thorough analysis of the comparative contributions of age of onset and frequency of use. We argue below that language dominance, instead of an incomplete grammar, likely plays the defining role in the findings demonstrated in this investigation.

4.2 Language activation

Investigations of bilingual processing using behavioral and neuro-physiological measures on lexical access tasks have shown that bilinguals activate both languages when only one is presented (Grosjean, 2001). While activation of both the HL and ML of the heritage speaker could be a source of differences in processing, it is not likely given that the comparison group was also bilingual and therefore, would also have dealt with simultaneous activation pressures. However, due to differences in dominance between the bilingual groups, activation patterns could be different. Heritage speakers might therefore have to be more efficient in modulating limited cognitive resources under pressures from non-dominant language processing and suppression of the activated dominant language.

Current models of bilingual lexical processing, specifically the Revised Hierarchical Model (RHM, cf. Kroll & Stewart, 1994), account for parallel activation of both (or all) of a bilingual’s languages, even when an individual is receiving input in only one language. These findings were originally based on language ambiguous stimuli (e.g., cognates or interlingual homographs), or measures of the processing cost associated with language switching. Costa, La Heij, and Navarrete (2006) raised the point that using switching tasks to evaluate evidence for parallel activation of a bilingual’s two languages leads to an unnatural language production context. Grosjean’s (2001) language mode hypothesis suggests that the language context of an experiment, whether it is set implicitly or explicitly, can influence bilingual functioning by manipulating the participant’s activation along the “monolingual–bilingual continuum” of
language use. Concern regarding the effects of unnatural language contexts in experimental studies of language processing was explored by Wu and Thierry (2010). Wu and Thierry argue that theoretical accounts of language control must take the language context of the experiment into account. To evaluate evidence for parallel activation in a monolingual mode, researchers should restrict experiments and experimental settings to a single language and offer guidelines to eliminate potentially confounding variables associated with language context.

Putnam and Sanchez (2013) argue that the patterns of HL and ML activation specific to heritage speakers influence the composition of lexical items, formal linguistic features, and parsing strategies in the HL. They argue that their model can account for the various outcomes in production and perception observed in heritage speakers’ HL skills, including the phenomenon of receptive bilinguals. This is because the model concentrates on activation (language that is produced and attended to, not simply present in the ambient environment). They further argue that some late-acquired HL features may not have received sufficient prior activation to compete with the escalated activation of the societally dominant language, formal features associated with the HL may instead resemble the ML.

Following Kroll and Stewart’s (1994) Revised Hierarchy Model, the relationships between semantic concepts and access to matching lexical items in each language can change as an individual’s proficiency in each language changes. Switching language dominance can have an impact on the strength and directionality of the connections between lexical items, where the lexical items from the second acquired language can not only develop a direct connection to a concept but can also come to block the activation of the first acquired lexical item for that concept. Additionally, with increased frequency of use of the ML and reduced frequency of use of the HL, the lexical items of the HL become harder to access, as per Paradis’s (1993, 1996) Activation Threshold Hypothesis. Because the HL faces substantial mental suppression/inhibition, the amount of effort needed to activate lexical items and formal features of the home language increases. This pattern increases over the years that the heritage speaker has reduced use of the HL making it increasingly difficult to access HL features as time progresses.

The suppression/inhibition interacts with language processing in the context of individual natural cognitive limitations. If cognitive effort is being devoted to suppression/inhibition of the ML when processing the HL, the constraints on cognitive resources could results in underperformance on
“resource-intensive phenomena like binding ... or scope inversion” as Scontras, Fuchs, and Polinsky (2015, p. 5) conjecture about their findings and the findings from Kim (2007). In the present study, the cognitive resources already dedicated to suppression/inhibition of the ML limit the available resources to parse relative clauses. This necessitates more efficient parsing strategies resulting in the failure to have a subject-object relative clause processing asymmetry.

4.3 Heritage Language Anxiety

The increased processing load for subject-relatives observed in the present study with psychophysiological measures is consistent with the effects of anxiety, namely heritage language anxiety. In the ERP study, we observed a lack of early P300-like and centroparietal P600 effects for object-relatives and greater frontal and centroparietal P600 amplitude for subject-relatives. This attenuation of early ERP components and greater amplitude in later ERP components mirrors the effects of individuals with high math anxiety.

Math anxiety, which is distinct from general trait anxiety or test anxiety, is anxiety related to math or number-related tasks (Chang & Beilock, 2016). High math anxiety individuals have attenuated early ERP amplitude, such as P300 amplitude, that is not necessarily associated with poorer performance (Murray & Janelle, 2007) and is independent of task complexity or general trait anxiety levels (Klados et al., 2015). The attenuated early ERP components are also followed by greater amplitude for ERP components related to late-stage processing in high math anxiety individuals. The greater P600 amplitude in high math anxiety individuals (Suárez-Pellicioni, Núñez-Peña, & Colomé, 2013) may reflect up-regulation of cognitive control to mitigate negative performance as a result of math anxiety (Suárez-Pellicioni, Núñez-Peña, & Colomé, 2014). In high math anxiety individuals, poor math performance can also be avoided with increased frontoparietal activity (Chang et al., 2017) suggesting they are “ramping up control resources” (Lyons & Beilock, 2012). We can extend the findings of math anxiety research to the present study of heritage speakers to explain the observed lack of P300-like effects in object-relatives and increased P600 amplitude and in subject-relatives. The lack of ERP effects would normally implicate failure to parse the stimuli, however, math anxiety literature suggests that heritage speakers must be
utilizing different cognitive strategies to parse relative clauses consistent with the observed increased in task-evoked pupillary responses across condition.

We are not suggesting that heritage speakers have higher trait anxiety, being generally more anxious, or even high state anxiety, related to the experimental task, but rather, we are observing the effects of heritage language anxiety (Tallon, 2011). Heritage language anxiety – the anxiety associated with using a heritage minority language – was first proposed by Tallon (2011) in research focusing on heritage speakers learning the heritage language in a classroom setting. Heritage language anxiety is distinct from foreign language anxiety (Horwitz, Horwitz, & Cope, 1986) – the anxiety experienced by a learner of a foreign or second language in a classroom – as heritage speakers are not second or foreign language learners and heritage speakers are reported to have lower levels of foreign language anxiety than non-heritage learners (Tallon, 2011; Luo, 2015). Heritage language anxiety is also distinct from task-specific anxiety, such as writing, reading, speaking, or test-taking anxiety (Xiao & Wong, 2014). Crucially, heritage language anxiety is distinct from (meta-)linguistic insecurity – the lack of confidence in reporting metalinguistic evaluative judgements – as discussed in sections 4.2 - 4.3 of chapter II. (Meta-)Linguistic insecurity was specifically controlled for in the present study as the ERP and pupillometry protocols do not require specific metalinguistic judgments.

Heritage speakers report higher levels of anxiety related to using the heritage language (heritage language anxiety) with family and friends from the home country and related to using the majority language (majority language anxiety) around monolingual majority language speakers (Sevinç & Dewaele, 2018). This anxiety around use of the heritage language has a negative feedback loop with proficiency: the more anxious you feel about using the heritage language the less you use it, the less proficient you are and the more anxious you feel about using your lower proficiency heritage language (Sevinç & Backus, 2017). Sevinç (2017) utilized non-neural, electrodermal, implicit measures of anxiety (i.e., skin conductance levels, skin conductance response) in a video story re-telling task in the heritage language. Higher physiological anxiety responses were observed for heritage speakers with earlier age of acquisition of the majority language, lower oral proficiency in the heritage language and lower frequency of heritage language daily. Heritage language anxiety responses were negatively correlated with majority language anxiety responses.
Heritage language anxiety in the present study may have occurred in the heritage speakers as the task is in Spanish, their heritage language. Pupillometry, similar to galvanic skin response measures, indexes activation of the autonomic nervous system. Therefore, the large pupillary responses in both subject- and object-relative conditions could be confounded or masked by autonomic responses simultaneously indexing heritage language anxiety. Furthermore, heritage language anxiety likely has similar effects on ERPs as math anxiety. The findings of the present study suggest that we are observing heritage language anxiety which then must be accounted for in a model of language processing for heritage.

5. Efficient resource allocation

While object-relatives are harder to process, evidence of a subject-object relative clause processing asymmetry would not be found in heritage speakers if they used more efficient processing strategies. Processing efficiency as indexed by reduced cognitive effort has previously been reported across a number of domains. When performing mathematical tasks, individuals with higher intelligence show less pupil dilation as the task difficult increases (Ahern & Beatty, 1979). Lee and colleagues (2015) found that individuals with higher intelligence more efficiently modulated resource allocation as a function of visuo-spatial and standardized-test style language related task demands. Fewer resources, as indexed by less pupil dilation and more eye blinks\(^5\) were allocated to easier tasks and more resources, as indexed by greater pupil dilation and less eye blinks, were allocated to harder tasks. Individuals with lower intelligence, however, allocated the same amount of resources regardless of the task difficulty. A similar pattern of greater resource allocation only for the most difficult tasks was also found in complex geometric analogy tasks (van der Meer et al., 2010). Likewise, Piquado, Isaacowitz, and Wingfield (2010) found that older adults only showed greater processing load for object-relatives compared to subject-relatives in more complex sentences, reflecting more efficiency in processing. This strongly suggests that resource allocation is not merely a function of task difficulty but can be regulated by individuals who are more efficient at processing, such as the heritage speakers in the present study.

\(^5\) Eyeblinks were recorded in the in both the ERP and TEPR data but are removed as artifacts, as is standard protocol in data cleaning. The number of blinks could be analyzed as a measure of resource allocation in future studies. However, eye blink analysis is outside the scope of the present study.
We argue that heritage speakers are more efficient at processing due to limited cognitive capacity available to be allocated to relative clause processing as a result of the confounds additional cognitive processes introduced in section four. Any model of heritage speaker processing needs to consider the additional factors of heritage language activation, majority language suppression, and heritage language anxiety. These additional cognitive demands are evidenced by greater ERP and pupillary responses for subject-relatives in heritage speakers. As demonstrated by the present study, heritage speakers are able to comprehend relative clauses, therefore, the lack of an observed asymmetry is not explainable by positing an incomplete grammar. In this section we propose a model of heritage speaker processing that is efficient given additional cognitive constraints and is motivated by research in multitasking while addressing the negative impact of anxiety.

Assuming finite cognitive capacity in all speakers and to successfully process relative clauses, one must syntactically integrate the filler in subject-relatives, syntactically reanalyzing object-relatives and hold fillers in working memory in object-relatives, heritage speaker must utilize more limited cognitive capacity to complete these tasks. As schematized in Figure 58, finite cognitive capacity is represented as fixed for all speakers in all conditions. Cognitive resources associated with activation of the target language are different across speaker type. In monolinguals, activation of their only language is associate with limited resource consumption. In bilinguals, activation of the target language coöccurs with suppression of the other language they use and know. In late bilinguals, fewer costs are associated with activation of their first language, whereas, in heritage speakers, more cognitive resources are employed in the activation of the less-dominant heritage language. Likewise, in late bilinguals, suppression of their non-dominant L2, English, requires fewer cognitive resources than suppression of English, their dominant language, while this is not the case for heritage speakers. Lastly, heritage speakers have the additional burden of cognitive resources consumed by heritage language anxiety-related cognitive resource up-regulation in order to mitigate anxiety-related changes in performance. This results in differing amounts of remaining cognitive capacity that can be utilized in relative clause processing. Late bilinguals and monolinguals likely then have ample un-utilized cognitive capacity, assuming the lack of the additional burden of cognitive resources consumed by language anxiety-related cognitive resource up-regulation, to perform the additional relative clause related tasks without necessitating more strategic processing.
strategies. However, for heritage speakers, the successful completion of additional cognitive tasks associated with relative clause processing with limited cognitive resources available requires that heritage speakers must use more efficient processing strategies.

Figure 58: Schema of cognitive resource capacity and utilization in heritage speakers, late bilinguals and monolinguals.

Cognitive control, the process that modulates non-automatic information processing (Schmidt, 2001) by the allocation of brain resources (Christie & Schrater, 2015), is limited and has a low capacity (Wu et al., 2016). Due to the limited capacity for the additional relative clause processing tasks, more efficient processing strategies must be utilized by heritage speakers. More efficient processing would circumvent the “central bottleneck” due to limited capacity and avoid the processing tasks being carried out sequentially (Welford, 1952; Pashler, 1984). The central bottleneck, normally discussed for attention demanding tasks, has been established as a limiting factor in seemingly automatic language processing as well (Ferreira & Pashler, 2002).

Our proposal of more efficient processing in heritage speakers is at odds with our inclusion of heritage language anxiety mitigation effects that in part result in the capacity limitations, as anxiety has a deleterious effect on cognitive control and efficiency. Anxiety has been argued to impair processing efficiency more than performance (Eysenck, 1996). The two major theories of anxiety effects in processing efficiency, the processing efficiency theory (Eysenck & Calvo, 1992) and the attentional control theory (Eysenck, Derakshan, Santos, & Calvo, 2007), argue that anxiety negatively impacts the
capacity and processing of working memory or cognitive control (central executive function), respectively (see Derakshan & Eysenck (2009) for a review of the two major theories). Some studies have found evidence that anxiety reduces working memory (e.g., Darke, 1988), whereas others have found that anxiety disrupts central executive processing (Ashcraft & Kirk, 2001). The reduction of cognitive control negatively impacts top-down goal-directed processes (Eysenck, Derakshan, Santos, & Calvo, 2007), such as inhibition, shifting, and updating functions (Miyake et al., 2000; Wong, Mahar, Titchener, & Freeman, 2013; Qi et al., 2014) resulting in a decline in processing efficiency. If heritage speakers are experiencing greater anxiety, then we would expect that their performance (comprehension of relative clauses) would not be impacted but they would be less efficient in their processing (more effort would be required). They also would not be able to utilize more efficient processing strategies as their cognitive control processes would be negatively impacted due to heritage language anxiety. However, it is important to note that heritage speakers differ from late bilinguals and monolinguals in that they are proficient bilinguals who have been using and switching between two or more languages since early childhood. We therefore suggest that the executive function benefits of early and consistent bilingualism mitigates any negative anxiety related executive function effects.

The lifelong frequent use of two languages has been argued to “exercise brain networks responsible for cognitive control” (Kroll, Bobb, & Hoshino, 2014). The enhancement of executive functions (Bialystok, 2001) has been demonstrated in children and adults and has been argued to offset decline in executive function processes due to senescence (Bialystok, Craig, Klein, & Viswanathan, 2004; Bialystok & Viswanathan, 2009; Qu et al., 2015). Neuroimaging studies have found increased activation of the anterior cingulate cortex which has been argued to reflect more efficient processing in young adult bilinguals (Abutalebi et al., 2012) and older adult bilinguals (Gold et al., 2013) relative to monolinguals. For a thorough review of the consensus view of the bilingual advantage, see Bialystok and colleagues (2009) and Bialystok (2015). Likewise, see Bialystok Craik, and Luk (2012) for a review of the replicated and confound-controlled studies showing high-level cognitive function benefits and delayed onset of dementia in bilingual adults. However, some researchers argue against a bilingual advantage to executive function processes (Valian, 2015), others have reported possible counter-evidence (e.g., Kousaie et al., 2014; Paap et al., 2016), critiqued the generalizability of the executive function benefits
(Paap & Sawi, 2014), or presented methodological critiques of the bilingual advantage findings (Calvo, García, Manoiloff, & Ibáñez, 2015; Yang, Hartanto, & Yang, 2016). Given that replications are exceedingly rare, we will take the position that the bilingual advantage findings which suggest that it mitigates the negative effects of anxiety in central cognitive control functions have merit. Likewise, Poarch and Bialystok (2015) found better bilingual performance relative to monolinguals on a flanker task which requires enhanced selection, inhibition, and shifting processes, the same processes that are negatively impacted by anxiety.

We conjecture that heritage speakers are utilizing domain-general multitasking strategies to maintain high performance (comprehension of the heritage language) by efficiently executing the cognitive tasks required by relative clause processing (working memory demands, syntactic reanalysis) in parallel. Multitasking refers to the execution of multiple tasks utilizing the same processes simultaneously, and has been observed in language processing (Bailer & Tomitch, 2016). Multitasking typically is associated with an inverted U curve with productivity: medium multitasking leads to greater productivity than low multitasking or high multitasking. In terms of accuracy, multitasking is associated with a decline in accuracy (Adler & Benbunan-Fich, 2012). These effects are dependent on the difficulty of the tasks being performed where better performance is observed with simpler tasks (Adler & Benbunan-Fich, 2015). The decline in performance is argued to result from the reduced ability to control cross-talk between the processes being executed as the number of tasks performed in parallel increase (Feng, Schwemmer, Gershman, & Cohen, 2014). There was no decline in performance, as heritage speakers performed at ceiling on comprehension tasks. We therefore suggest that they are efficient multitaskers. Efficient multitaskers are able to shift between more efficient parallel and resource-saving serial processing depending on task or state (e.g., under stress) (Fischer & Plessow, 2015). Serial processing of object-relatives in the present study would result in delayed ERP components, which were not explored in the present study, and later pupil dilation effects, which were not observed. This suggests that heritage speakers are executing the object-relative processing tasks in parallel. Multitasking is not always associated with a decline in performance. For some people (2.5% of the population), there are no costs associated with multitasking (Watson & Strayer, 2010). Neuroimaging studies of efficient multitaskers (“extraordinary multitaskers” or “supertaskers”) found less activation in the anterior cingulate and posterior
frontopolar prefrontal cortices suggesting more efficient utilization of limited resources to effectively manage additional task goals and cognitive load (Medeiros-Ward, Watson & Strayer, 2014). We are not suggesting that heritage speakers are all supertaskers by virtue of some innate predisposition, but rather that efficient multitasking is practiced and acquired as a result of their lifelong bilingualism. Multitasking ability is mutable and not a fixed trait of one’s cognitive ability, but rather can be trained (Dux et al., 2009; Bender et al., 2017) and even forced. A recent US military study was able to “create” efficient multitaskers who were able to processes more information more quickly using transcranial direct current stimulation to increase neuronal excitability in the left dorsolateral prefrontal cortex (Nelson et al., 2016). An obvious counterargument against the proposal that heritage speakers are efficient multitaskers or supertaskers is that there is no evidence, even anecdotal evidence, that heritage speakers are outperform monolinguals in the workplace or school. We suggest that the efficient multitasking proposal is limited to language processing and not a domain-general cognitive strategy, following Bender and colleagues (2017), which found that multitasking training improved task-specific performance but didn’t extend to untrained tasks.

An additional caveat is warranted. We are also not suggesting that efficient multitasking of language processing is an inherent property of all heritage speakers or a fixed cognitive strategy. It is likely that efficient multitasking could change as a function of the language behaviors of a heritage speaker over their lifetime. The heritage language could become easier to access in the case of greater immersion (Linck, Kroll, & Sunderman, 2009) such as if the heritage speaker moved into a majority heritage language speaking neighborhood or worked in a heritage language speaking workplace. Likewise, the heritage language could become associated with fewer processing costs if there were changes in the frequency of codeswitching (Beatty-Martínez & Dussais, 2017) that could occur if heritage language use patterns changed with more or less ethnolinguistic cohort interaction. If processing costs of the heritage language decreased or heritage language anxiety decreased, efficient multitasking would not be needed. The present study included heritage speakers who were dominant in the majority language, and therefore efficient multitasking was required.

In conclusion, we propose a model of language processing for heritage speakers that accounts for successful comprehension and utilization of the heritage language and accounts for the lack of observed subject-object processing asymmetry effects. The subject-object processing asymmetry is not
subject to cross-linguistic influence due to opposing or conflicting syntactic representations in either language of the heritage speaker. We argue that the lack of observed processing effects of linguistic structures that heritage speakers parse successfully result from the additional cognitive costs associated with activation of the non-dominant heritage language, suppression of the dominant majority language, and heritage language anxiety mitigation processes. We argue that the deleterious effects of heritage language anxiety on cognitive control are mitigated by the cognitive control enhancements of lifelong bilingualism and frequent use of two languages. As a consequence of the limited cognitive resources available to syntactically parse relative clauses, we propose that heritage speakers must utilize efficient multitasking processing strategies similar to supertaskers.

6. Recommendations

In order to effectively study heritage speakers and de-center the monolingual in heritage speaker research, certain considerations must be made. Echoing the argument made by Kroll, Dussias, Bogulski, and Valdes Kroff (2012), there is more to learn about cognition and language in the study of multilingualism, including heritage speakers, than in studying monolinguals. In previous chapters we have motivated the need to appropriate comparison groups, appropriate research paradigms, and appropriate linguistic phenomenon. The contributions of this dissertation are the motivation of the importance of accounting for language anxiety in processing studies and the proposal of a model of bilingual language processing using efficient processing strategies. Using these insights, we are able to make additional recommendations for the study of language processing, namely, the need for more attention to stimuli design, the pupillometry as an effective measure of anxiety, and the need appropriate statistical techniques to control for language dominance.

6.1 Anxiety screening

Language anxiety has been shown to be a confound that has implications not just for study of heritage speakers, but all speakers. Different populations of language users can have anxiety around language that should be controlled for. Language anxiety in heritage speakers is possible when testing/using the heritage language as heritage language anxiety and when testing/using the social
majority language as majority language anxiety. Language anxiety in late bilinguals is possible when testing/using the L2 as majority language anxiety and when testing/using the L1 as anxiety resulting from anti-immigrant or anti-bilingualism bigotry. Language anxiety in late learners is possible when testing/using the L2 as foreign language anxiety. Even in monolinguals, language anxiety is possible when testing/using the only language they command. Speakers of socially-devalued and discriminated language varieties, such as varieties of a disadvantaged socioeconomic or ethno-racial community, anecdotally refer to their variety as “bad”, “improper”, or “uneducated”. The outright prejudice of the dominant social group, whose language is rarefied and codified as the “proper”, “educated”, and “standard” variety, negatively impact monolinguals feelings about their language. The thinly-veiled elitist, racist, and classist propagation of prescriptivist notions of “good” grammar likely provokes language anxiety in monolinguals. Furthermore, language anxiety is likely also present in clinical populations, such as individuals with specific language impairment, dementia, downs syndrome, parkinson’s disease, etc.

In addition to language anxiety, anxiety resulting from being tested in a laboratory setting by strangers is possible. Being asked to perform language through comprehension tasks, story (re-)telling tasks, or metalinguistic acceptability judgement tasks, etc. likely result in higher levels of anxiety in participants. Likewise, the unnatural act of being tested wearing an electrode cap in ERP studies, or lying in an MRI machine, or sitting in a soundproof booth likely also provoke higher levels of anxiety. Similarly, the L1 of the researchers running the study might provoke higher levels of anxiety. For example, in a situation where the researchers are L1 English speakers and interacting with an L2 English-speaking participant in English, a participant might be more anxious. In the study of anxiety, a distinction is made between trait anxiety, a persistent personality characteristic of individuals, and state anxiety, a temporary anxiety response. Language and task related anxiety are transitory anxiety responses to a specific situation and not enduring characteristics of people’s personality. It is crucial that state anxiety be specifically measured or accounted for in studies of language processing. Trait anxiety is also a confound that should be specifically incorporated into experimental studies of language. Individuals who have more anxious personalities confound the results in a pupillometry or ERP study.

Therefore, we recommend that well-establish psychometric measures of anxiety be incorporated as screeners or controls in experimental language studies. Furthermore, experimental language studies
must also include measures of language anxiety and task anxiety to control for anxiety-related effects. As there are limited validated measures of language anxiety, we suggest that researchers develop assessments or questionnaires as a first step in accounting for state anxiety. State anxiety resulting from the experimental tasks is a confound for all participants so in a sense it is ever-present noise that impacts participants equally. However, the effects of task anxiety on ERP or pupillometry responses should be taken into consideration when interpreting results.

6.1 Stimuli design

Stimuli must be constructed in such a way to control for the emotionality of the lexical items. Individuals with higher levels of (language) anxiety, are unable to down-regulate emotional responses as the brain reacts more rapidly and vigorously (Eldar et al., 2010). The unregulated effects of emotional word processing in high anxiety individuals could then introduce noise, confounding ERP study results when comparing their ERP responses to non-anxious groups. If the group distinction made by the experimenter overlaps with anxiety (e.g., group 1 has more anxiety around use of the prestige variety, group 2 has less anxiety around the use of the prestige variety) then group differences found by the researchers are not able to reliably be attributed to the effect of group difference. Researchers have found that emotional word (positive, neutral, negative) processing results greater P300/P400/P600 amplitude in monolinguals (Herbert et al., 2006; Zhang et al., 2014) and in bilinguals in both the L1 and L2 (Conrad, Recio, & Jacobs, 2011). This effect is greater for individuals with higher physiological symptoms of anxiety (Sehlmeyer et al., 2010) or had higher trait anxiety (Weinstein, 1995). Conversely, some researchers have found attenuated ERP component amplitude as a result of anxiety. Emotional word processing in high anxiety individuals has been found to result in smaller N2 amplitude as well as changes in spatial distribution of ERP components (de Pascalis & Morelli, 1990). Some researchers argue that the attenuation of ERP components reflecting late-stage processing, such as the P300, indexes the favoring of bottom-up processing as a function of arousal/distress (Rossi & Pourtois, 2017). Evidence of emotionality introducing noise in pupillometry comes from Harris, Ayçiçeği, and Gleason’s (2003) study of galvanic skin responses to the emotionality of words in the L1 and L2 of bilinguals. They found that autonomic responses were greater to seemingly innocuous negative words in the L1 (e.g., crime, fight,
sick) and the autonomic responses were stronger with auditory presentation. These studies suggest that emotional word processing, which is happens concurrently with syntacto-semantic processing introduces noise in any study of language processing. As the ERP component and pupillometry effects are modulated by non-task related anxiety (task-related anxiety would be the same for all participants in a group and therefore assumed to be constant), high anxiety individuals’ syntacto-semantic processing as measured by ERP components and pupillometry effects are confounded. The results are obscured by emotional word processing ERP component and pupillometry effects. We recommend that in addition to controlling for word frequency, researchers control for the emotionality of words in the sentence stimuli.

6.2 Pupillometry as a measure of anxiety

While pupillometry initially appeared to be an ideal measure given that it is an implicit measure with limited set-up (no need for the 45 minutes to fit an ERP cap), it is unfortunately confounded by anxiety. Autonomic activation resulting from anxiety potentially mask any cognitive load related responses. Pupillometry is too crude a measure in that it is the sum of a number of cognitive processes and psychological states that cannot be disentangled in the same way as ERP measures. The ERP effects of anxiety are well documented, so anxiety related effects can be mitigated or accounted for in a way. Therefore, as a measure of language processing, pupillometry is unreliable. Numerous studies have found task-related pupil responses to language processing. We are not suggesting that those findings are spurious, only that a null result in a pupillometry is uninterpretable. If task-related pupil responses to language processing for specific experimental conditions are not observed, the researcher is unable to determine why the participants failed to show evidence of differential processing load to the conditions. It could reflect processing differences, or it could possibly reflect anxiety. Unless anxiety can be controlled for in pupillometry studies, it is an ineffective protocol to use in studying language processing and grammatical knowledge. However, as pupillometry is a measure of autonomic activation, it is an ideal measure for directly measuring anxiety. We suggest that future studies use pupillometry in the study of language anxiety. Simultaneous pupillometry and ERP recordings could be used to determine anxiety-levels during ERP tasks to reduce the impact of anxiety on ERP studies/findings.
6.3 Control for anxiety in ERP studies

Event-related potential studies are not immune to the effects of language anxiety, however, given the body of literature on anxiety and ERP, researchers can still utilize ERP as long as appropriate statistical techniques are incorporated. Language anxiety, state/task anxiety, and trait anxiety should be incorporated into statistical models of ERP component amplitude. The inclusion of variables which operationalize anxiety would be able to then reveal ERP component effects in regression models. Statistical techniques, such as t-tests or ANOVAs which do not allow for the explicit modeling of anxiety would fail to disentangle anxiety-related attenuation and amplification of ERP components. Recent studies by Martohardjono and colleagues (2016, 2017) have found that when modeling ERP component amplitude with bilingual group alone, no effects were found in heritage speakers, but when modeling the ERP component amplitude with variables indexing language use and exposure (which can be argued to index heritage language anxiety), significant ERP effects were observed. It is therefore crucial that researchers incorporate measures of anxiety into any study of processing. As mentioned above, linguistic anxiety is not unique to heritage speakers. Late bilinguals, language learners, and even monolinguals are susceptible to language anxiety effects that would need to be modeled and controlled for.

7. Future studies

The present study highlights the effect of language dominance in bilinguals. We explained the findings of the present study by arguing that heritage speaker dominance in the majority language, English, when using their heritage language, Spanish, is a confound when studying processing. Therefore, we predict that when testing heritage speakers in the majority language and late bilinguals in their L2, the results would be flipped. We expect that a subject-object relative clause processing asymmetry would then be observable in heritage speakers due to reduced costs associated with the activation of their dominant language, suppression of their non-dominant language, and absence of heritage language anxiety, although there might be autonomic activation as a results of majority language anxiety.

Furthermore, to explore the effect of language dominance, an analysis incorporating individual language exposure, language use, and language ability will be conducted similar to Martohardjono and colleagues (2016, 2017). The sociolinguistic variables will be derived from the language background
questionnaire reported in section 3 of chapter II. The factors from the questionnaire will then be used in mixed-effects modeling without group to explore the effect of language use and exposure in predicting either ERP or TEPR amplitude.

Event-related potential results inform us about the different stages of processing after encountering the gap (in object-relatives) or filled gap (in subject-relatives). We are interested in further exploring the processing differences as a function of time. In order to do this, we intend to track the eye gaze of bilinguals while processing relative clauses in a picture selection task to better understand the stages of processing and comprehension in these two different groups of bilinguals.

As mentioned previously in section 7.4 of chapter IV, a PCA was used to isolate the linguistic ERP components related to sentence processing. The analysis was performed across bilingual group which assumes that temporal and spatial characteristics of the ERP component responses are similar across group. A number of studies have found latency and spatial differences between bilingual groups (e.g., Kotz et al., 2008; Moreno & Kutas, 2005) and even monolingual groups differing on standardized measures of proficiency (e.g., Weber et al., 2003; Pakulak & Neville, 2010). Therefore future studies will run the temporo-spatial PCA by bilingual group to characterize the early and late wave topographic and latency differences between heritage speakers and late bilinguals.

The items in the ERP and TEPR studies were object-embedded meaning that the relativized noun was the object of the matrix clause. The majority of psycholinguistic studies utilize stimuli where the head noun is a subject in the matrix clause. One of the many theorized sources of the subject-object relative clause processing asymmetry is structural or thematic mismatch (Traxler, Morris, & Seely, 2002). In the present study, difficulty arising from structural or thematic mismatch are different for object-embedded relative clauses than for subject-embedded relative clauses. The head noun in the current study was the matrix object which could have resulted in the subject-relative being harder to process due to the mismatch between the head noun’s thematic or structural role in the matrix clause and the subordinate clause. In an ERP study, thematic reanalysis has been found to elicit a posterior positivity 300-600ms post-event, a P300/P400 (Bomkessel et al., 2003). While no present studies to have explored the effect of thematic mismatch using pupillometry, the P300 has been linked to locus coeruleus-norepinephrine system activation (Nieuwenhuis, de Geus, & Aston-Jones, 2013) which can be measured
as increased pupil size. A future study is needed which examines the effect of subject- versus object-embedded sentences to see if there are differences in sensitivity to thematic or structural mismatch between heritage speakers and late bilinguals.
APPENDIX A:
SGBP LANGUAGE BACKGROUND QUESTIONNAIRE

Demographics
Where were you born?
If born abroad, when did you arrive in the US?
How old are you?
Participant's profession in U.S.: ________________________________
Participant's social class (choose one): working ____ middle ____ upper____

Language Exposure
What language did you speak with your primary caregiver(s) from birth to age 10?
What was the first language you learned? ________________________
What languages were spoken in your house growing up? __________________________
Which of the languages from above were used most often? _________________________
Who spoke each of the languages in (14.) to each other in your house growing up?
Example: Language __Spanish____: everyone spoke Spanish to each other
Language __Nahuatl____: grandparents spoke Nahuatl to each other and no one else
Language ___________: _________________________________________________
Language ___________: _________________________________________________
Language ___________: _________________________________________________

Please complete the following table: (by year age 5-18)
What country did you live in?
What was the primary language spoken in your local community?
Did you attend school?
What was the language of instruction?

Language Ability & Identity
How well do you understand Spanish:
1 = little to nothing of what I hear
2 = some of what I hear
3 = about half of what I hear
4 = most of what I hear
5 = everything I hear

What do you consider to be your native language?
Please list all the languages that you speak (do not include languages that you can read but do not speak):

For level:  
1 = I have limited knowledge of the language  
2 = I have some ability to use the language  
3 = I have good ability to use the language*  
4 = I am a fluent speaker/user of the language  
5 = I am a native speaker/user of the language

*If you select “3 = I have good ability to use the language”, please write “YES” if you are able to give an opinion and defend it in that language.

Language _______________, level 1 2 3 4 5, when did you start learning? _____ years old
Language _______________, level 1 2 3 4 5, when did you start learning? _____ years old
Language _______________, level 1 2 3 4 5, when did you start learning? _____ years old
Language _______________, level 1 2 3 4 5, when did you start learning? _____ years old

Which languages do you read/write? At what level? When did you start?

For level:  
1 = I have limited reading/writing ability in the language  
2 = I have some ability to read/write in the language  
3 = I have good ability to read/write in the language*  
4 = I am a fluent reader/writer of the language  
5 = I am a native reader/writer of the language

*If you select “3 = I have good ability to read/write in the language”, please write “YES” if you are able to defend an opinion in writing in that language.

Language _______________, level 1 2 3 4 5, when did you start? _____ years old
Language _______________, level 1 2 3 4 5, when did you start? _____ years old
Language _______________, level 1 2 3 4 5, when did you start? _____ years old
Language _______________, level 1 2 3 4 5, when did you start? _____ years old

Language Use

Which language(s) do you use to speak with your:

father .............................. English / Spanish / both / N/A
mother .............................. English / Spanish / both / N/A
sisters/brothers .................... English / Spanish / both / N/A
children (older) ...................... English / Spanish / both / N/A
children (younger) .................. English / Spanish / both / N/A
friends ............................. English / Spanish / both / N/A
boss .................................. English / Spanish / both / N/A
co-workers .......................... English / Spanish / both / N/A
classmates ........................... English / Spanish / both / N/A
significant other .................... English / Spanish / both / N/A

How much Spanish do you use in/at:

home .............................. mostly / little / none / N/A
school .............................. mostly / little / none / N/A
work .............................. mostly / little / none / N/A
social activities .................... mostly / little / none / N/A
reading ............................ mostly / little / none / N/A
listening to the radio/music ..... mostly / little / none / N/A
watching TV ........................ mostly / little / none / N/A
APPENDIX B:  
SGBP ACCEPTABILITY JUDGMENT TASK STIMULI

Complex noun phrase complement islands with object wh-extraction (CCOE)

Grammatical
1. El jefe confesó el hecho que la agencia eliminó el sindicato.  
   ¿Qué sindicato confesó el jefe que la agencia eliminó?
2. La periodista reportó la noticia que el equipo perdió el partido.  
   ¿Qué partido reportó la periodista que el equipo perdió?
3. El estudiante escribió el reporte que la reacción produjo un carbohidrato complejo.  
   ¿Qué carbohidrato complejo escribió el estudiante que la reacción produjo?
4. Juan contó el chisme que el vecino robó un carro anoche.  
   ¿Qué carro contó Juan que el vecino robó anoche?
5. María creyó el rumor que el estudiante gastó el dinero.  
   ¿Qué dinero creyó María que el estudiante gastó?

Ungrammatical
1. Cecilia propuso la hipótesis que un evento cósmico formó la galaxia.  
   *¿Qué galaxia propuso Cecilia la hipótesis que un evento cósmico formó?
2. Gimena escuchó el reportaje que un insecto invade el campo.  
   *¿Qué insecto escuchó Gimena que invade el campo?
3. El cocinero negó la acusación que el dueño cerrará el restaurante.  
   *¿Qué restaurante negó el cocinero que cerrará?
4. El gobierno afirmó la creencia que el ejército produce un arma biológica.  
   *¿Qué arma biológica afirmó el gobierno que produce?
5. Ana oyó la noticia que el equipo ganó el partido de tenis.  
   *¿Qué partido de tenis oyó Ana la noticia que el equipo ganó?

Complex noun phrase complement islands with subject wh-extraction (CCSE)

Grammatical
1. Cecilia propuso la hipótesis que un evento cósmico formó la galaxia.  
   ¿Qué evento cósmico propuso Cecilia que formó la galaxia?
2. Gimena escuchó el reportaje que un insecto invade el campo.  
   ¿Qué insecto escuchó Gimena que invade el campo?
3. El cocinero negó la acusación que el dueño cerrará el restaurante.  
   ¿Qué dueño negó el cocinero que cerrará el restaurante?
4. El gobierno afirmó la creencia que el ejército produce un arma biológica.  
   ¿Qué ejército afirmó el gobierno que produce un arma biológica?
5. Ana oyó la noticia que el equipo ganó el partido de tenis.
¿Qué equipo oyó Ana que ganó el partido de tenis?

**Ungrammatical**

1. El veterinario explicó su teoría que la química afecta el olfato.
   *¿Qué química explicó el veterinario su teoría que afecta el olfato?

2. El representante leyó el anuncio que el ciudadano ganó la demanda.
   *¿Qué ciudadano leyó el representante el anuncio que ganó la demanda?

3. El inversionista rechaza el informe que su error impactó el mercado.
   *¿Qué error rechaza el inversionista el informe que impactó el mercado?

4. El naufrago mantiene la esperanza que la nave verá su señal.
   *¿Qué nave mantiene el naufrago la esperanza que verá su señal?

5. Pablo notó el hecho que el mecánico arregló el carro.
   *¿Qué mecánico notó Pablo el hecho que arregló el carro?

**Object-object relative clause island with subject wh-extraction (RCOO)**

**Grammatical**

1. El cine mostró el documental que el crítico odiaba.
   ¿Qué cine mostró el documental que el crítico odiaba?

2. La abogada citó la evidencia que el defensor negó.
   ¿Qué abogada citó la evidencia que el defensor negó?

3. El profesor tomó el café que el estudiante trajo.
   ¿Qué profesor tomó el café que el estudiante trajo?

4. El gato agarró el jamón que el perro comía.
   ¿Qué gato agarró el jamón que el perro comía?

5. El periódico publicó la noticia que el redactor escribió.
   ¿Qué periódico publicó la noticia que el redactor escribió?

**Ungrammatical**

1. El restaurante compró la sopa que la gente comió.
   *¿Qué gente compró el restaurante la sopa que comió?

2. La facción detestaba la maniobra que el político ejecutaba.
   *¿Qué político detestaba la facción la maniobra que ejecutaba?

3. El colegio diseñó el código que la administración aprobó.
   *¿Qué administración diseñó el colegio el código que aprobó?

4. El museo cerró la exposición que el gobierno financiaba.
   *¿Qué gobierno cerró el museo la exposición que financiaba?

5. La tienda vendía la pastilla que el enfermo tomó.
   *¿Qué enfermo vendía la tienda la pastilla que tomó?

**Object-subject relative clause island with object wh-extraction (RCOS)**

**Grammatical**

1. Josefina dejó la caja que contenía el maquillaje.
1. ¿Qué caja dejó Josefin que contenía el maquillaje?

2. Paola hizo el gesto que causó la controversia.
   ¿Qué gesto hizo Paola que causó la controversia?

3. El juez aprobó la decisión que modificó la ley.
   ¿Qué decisión aprobó el juez que modificó la ley?

4. Juan desarrolló el programa que llevaba la nómina.
   ¿Qué programa desarrolló Juan que llevaba la nómina?

5. Natalia conoce la empresa que pintó la casa.
   ¿Qué empresa conoce Natalia que pintó la casa?

**Ungrammatical**

1. Nico pintó el cuadro que estropeó la pared.
   *¿Qué pared pintó Nico el cuadro que estropeó?

2. Ignacio recomendó el restaurante que ganó el premio.
   *¿Qué premio recomendó Ignacio el restaurante que ganó?

3. María contrató el servicio que mandó el paquete.
   *¿Qué paquete contrató María el servicio que mandó?

4. La mujer tocó la música que llenaba la habitación.
   *¿Qué habitación tocó la mujer la música que llenaba?

5. Laura llevaba la bolsa que tenía el zapato.
   *¿Qué zapato llevaba Laura la bolsa que tenía?

**Subject relative clause island with object wh-extraction (RCSS)**

**Grammatical**

1. La pintora que recibió el premio dibujó un paisaje.
   ¿Qué paisaje dibujó la pintora que recibió el premio?

2. El periodista que ganó el premio escribió un libro.
   ¿Qué libro escribió el periodista que ganó el premio?

3. La investigadora que descubrió la molécula explicó su método.
   ¿Qué método explicó la investigadora que descubrió la molécula?

4. El conductor que chocó el colectivo cometió una infracción.
   ¿Qué infracción cometió el conductor que chocó el colectivo?

5. El gerente que maneja el negocio firmó la carta.
   ¿Qué carta firmó el gerente que maneja el negocio?

**Ungrammatical**

1. La lavandería que perdió la camisa ofreció un descuento.
   *¿Qué camisa la lavandería que perdió ofreció un descuento?

2. La trabajadora que recibió el aumento mejoró su producción.
   *¿Qué aumento la trabajadora que recibió mejoró su producción?

3. El embajador que manejó la negociación llevó la reunión.
*¿Qué negociación el embajador que manejó llevó la reunión?

4. El hombre que contestó el teléfono vio el concierto.
   *¿Qué teléfono el hombre que contestó vio el concierto?

5. El estudiante que tiró el lápiz copió el examen.
   *¿Qué lápiz el estudiante que tiró copió el examen?

Temporal adverbial island with object wh-extraction (TAOE)

Grammatical
1. El soldado cargó el rifle después que el general comenzó la misión.
   ¿Qué rifle cargó el soldado después que el general comenzó la misión?

2. El marinero manejó el remo mientras que el piloto guiaba el barco.
   ¿Qué remo manejó el marinero mientras que el piloto guiaba el barco?

3. El cocinero prendió el horno mientras que el mesero gritaba el pedido.
   ¿Qué horno prendió el cocinero mientras que el mesero gritaba el pedido?

4. El coronel firmó la declaración mientras que el ejército preparaba el bombardeo.
   ¿Qué declaración firmó el coronel mientras que el ejército preparaba el bombardeo?

5. El gato tumbó el jarrón después que el perro mordió el zapato.
   ¿Qué jarrón tumbó el gato después que el perro mordió el zapato?

Ungrammatical
1. El niño comió el dulce mientras que su tía buscaba la comida.
   *¿Qué comida comió el niño el dulce mientras que su tía buscaba?

2. El profesor leyó la nota mientras que el decano daba el discurso.
   *¿Qué discurso leyó el profesor la nota mientras que el decano daba?

3. La senadora aprobó la legislación mientras que el asistente grababa el evento.
   *¿Qué evento aprobó la senadora la legislación mientras que el asistente grababa?

4. El mesero trajo el trapo después que el hombre viró el plato.
   *¿Qué plato trajo el mesero el trapo después que el hombre viró?

5. El carro tumbó el árbol después que la camioneta golpeó el ciervo.
   *¿Qué ciervo tumbó el carro el árbol después que la camioneta golpeó?

Temporal adverbial island with subject wh-extraction (TASE)

Grammatical
1. El niño comió el dulce mientras que su tía buscaba la comida.
   ¿Qué niño comió el dulce mientras que su tía buscaba la comida?

2. El profesor leyó la nota mientras que el decano daba el discurso.
   ¿Qué profesor leyó la nota mientras que el decano daba el discurso?

3. La senadora aprobó la legislación mientras que el asistente grababa el evento.
   ¿Qué senadora aprobó la legislación mientras que el asistente grababa el evento?

4. El mesero trajo el trapo después que el hombre viró el plato.
   ¿Qué mesero trajo el trapo después que el hombre viró el plato?
5. El carro tumbó el árbol después que la camioneta golpeó el ciervo. ¿Qué carro tumbó el árbol después que la camioneta golpeó el ciervo?

Ungrammatical
1. La jueza dio el veredicto después que el defensor presentó su caso. *¿Qué defensor la jueza dio el veredicto después que presentó su caso?
2. El jardinero sacó la planta después que el veneno disolvió la raíz. *¿Qué veneno el jardinero sacó la planta después que disolvió la raíz?
3. El comité nombró un miembro nuevo después que la empresa confirmó la expansión. *¿Qué empresa el comité nombró un miembro nuevo después que confirmó la expansión?
4. El agente revisó el pasaporte mientras que el viajero buscaba el pasaje. *¿Qué viajero el agente revisó el pasaporte mientras que buscaba el pasaje?
5. El conductor paró el carro mientras que el niño cruzaba la calle. *¿Qué niño el conductor paró el carro mientras que cruzaba la calle?

Complement clause island with subject wh-extraction (TTSE)

Grammatical
1. Inés confesó que su hermana había comido la tarta. ¿Qué hermana confesó Inés que había comido la tarta?
2. Luis cree que el electricista cortó el cable. ¿Qué electricista cree Luis que cortó el cable?
3. Isabel creía a que la oficina había mandado el archivo. ¿Qué oficina creía Isabel que había mandado el archivo?
4. El cocinero sabe que el plato necesita un ingrediente. ¿Qué plato sabe el cocinero que necesita un ingrediente?
5. Laura dijo que el pollo arruinó la cena. ¿Qué pollo dijo Laura que arruinó la cena?
6. El agente reporta que la familia salió a las ocho. ¿Qué familia reporta el agente que salió a las ocho?
7. Hugo adivinó que el estudiante ensució el escritorio. ¿Qué estudiante adivinó Hugo que ensució el escritorio?

Ungrammatical
1. Tomás soñó que su perro tocaba guitarra. *¿Qué perro soñó Tomás tocaba guitarra?
2. Sofía jura que el banquero mandó la plata. *¿Qué banquero jura Sofía mandó la plata?
3. Carlos suponía que su hija había dejado el regalo. *¿Qué hija suponía Carlos había dejado el regalo?
4. El pronóstico indica que la tormenta llegará mañana. *¿Qué tormenta indica el pronóstico llegará mañana?
5. Olga pensó que la policía agarró a su hermano.
   *¿Qué policía pensó Olga agarró a su hermano?

6. Los científicos predicen que la población dejará de crecer.
   *¿Qué población predicen los científicos dejará de crecer?

7. El tribunal probó que la ley protege los derechos humanos.
   *¿Qué ley probó el tribunal protege los derechos humanos?

*Wh-island with object wh-extraction (WHOE)*

**Grammatical**

1. La policía teorizó dónde el terrorista había escondido la evidencia.
   *¿Qué evidencia teorizó la policía que el terrorista había escondido?

2. Olga dedujo dónde la hermana había escondido el dulce.
   ¿Qué dulce dedujo Olga que la hermana había escondido?

3. Martín sospechó por qué el niño había traído el correo.
   *¿Qué correo sospechó Martin que el niño había traído?

4. Ricardo explicó cómo la escuela remodeló el edificio.
   *¿Qué edificio explicó Ricardo que la escuela remodeló?

5. Ignacio admiraba cómo el niño resolvió la disputa.
   *¿Qué disputa admiraba Ignacio que el niño resolvió?

**Ungrammatical**

1. María cuestiona por qué el hermano necesitaba el carro.
   *¿Qué carro cuestiona María por qué el hermano necesitaba?

2. La secretaría adivinó cómo el empleado había saboteado el sistema.
   *¿Qué sistema adivinó la secretaría cómo el empleado había saboteado?

3. Juan odia cómo el profesor administra el examen.
   *¿Qué examen odia Juan cómo el profesor administra?

4. La madre descubrió por qué el niño había dejado el cuaderno.
   *¿Qué cuaderno descubrió la madre por qué el niño había dejado?

5. Jaime averiguó dónde el jubilado había tirado la comida.
   *¿Qué comida averiguó Jaime dónde el jubilado había tirado?

*Wh-island with subject wh-extraction (WHSE)*

**Grammatical**

1. María cuestiona por qué el hermano necesitaba el carro.
   ¿Qué hermano cuestiona María que necesitaba el carro?

2. La secretaría adivinó cómo el empleado había saboteado el sistema.
   ¿Qué empleado adivinó la secretaría cómo el empleado había saboteado el sistema?

3. Juan odia cómo el profesor administra el examen.
   ¿Qué profesor odia Juan que administra el examen?
4. La madre descubrió por qué el niño había dejado el cuaderno. ¿Qué niño descubrió la madre que había dejado el cuaderno?

5. Jaime averiguó dónde el jubilado había tirado la comida. ¿Qué jubilado averiguó Jaime que había tirado la comida?

**Ungrammatical**

1. Ignacio confirmó por qué la enfermera había llevado la medicina. *¿Qué enfermera confirmó Ignacio por qué había llevado la medicina?*

2. El médico estableció dónde el empleado había lastimado su dedo. *¿Qué empleado estableció el médico dónde había lastimado su dedo?*

3. Juan admitió dónde el hermano había alquilado la moto. *¿Qué hermano admitió Juan dónde había alquilado la moto?*

4. El vendedor notó cómo el muchacho robaba la mercancía. *¿Qué muchacho notó el vendedor cómo robaba la mercancía?*

5. El aficionado especuló por qué el entrenador vendaba la rodilla. *¿Qué entrenador especuló el aficionado por qué vendaba la rodilla?*
Subject-Object (sorc)

1. El perro, que el oso golpea, baila.
   *The dog, who the bear punches, dances*

   [CORRECT]

2. El mono, que el perro golpea, llora.
   *The monkey, who the dog punches, cries*

   [CORRECT]

3. El perro, que el oso abraza, salta.
   *The dog, who the bear hugs, jumps*

   [CORRECT]
4. El mono, que el perro patea, come.
*The monkey, who the dog kicks, eats*

```
Subject-Subject (ssrc)
```

1. El oso, que toca al perro, baila.
*The bear, who touches the dog, dances*

```
Subject-Subject (ssrc)
```

2. El conejo, que toca al oso, llora.
*The rabbit, who touches the bear, cries*
3. El conejo, que empuja al gato, baila.
The rabbit, who pushes the cat, dances

[CORRECT]

4. El perro, que abraza al mono, duerme.
The dog, who hugs the monkey, sleeps

[CORRECT]
APPENDIX D:
SGBP RELATIVE CLAUSE ITEMS

Subject relative clauses

1. Josefina dejó la caja que contenía el maquillaje.
   *Josefina left the box that contained the makeup*

2. Nico pintó el cuadro que estropeó la pared.
   *Nico painted the painting that ruined the wall*

3. Hugo organizó la fiesta que ensució el vecindario.
   *Hugo organized the party that made the neighborhood dirty*

4. Paola hizo el gesto que causó la controversia.
   *Paola made the joke that caused the controversy*

5. Ignacio recomendó el restaurante que ganó el premio.
   *Ignacio recommended the restaurant that won the prize*

6. Juan compró la cámara que tenía garantía.
   *Juan bought the camera that had a warranty*

7. El juez aprobó la decisión que modificó la ley.
   *The judge approved the ruling that modified the law*

8. María contrató el servicio que mandó el paquete.
   *Maria contracted the service that sent the package*

9. Sebastián usó el producto que limpia la computadora.
    *Sebastian used the product that cleaned the computer*

10. Juan desarrolló el programa que llevaba la nómina.
    *Juan developed the program that did payroll*

11. La mujer tocó la música que llenaba la habitación.
    *The woman played the music that was filling the room*

12. Jorge vio la película que ganó el premio.
    *Jorge saw the movie that won the award*

13. Natalia conoce la empresa que pintó la casa.
    *Natalia knew the company that painted the house*

14. Laura llevaba la bolsa que tenía el zapato.
    *Laura carried the bag that had the shoe*

15. Guillermo encontró la tienda que vendía la camisa.
    *Guillermo found the store that sold the shirt*
**Object relative clauses**

1. El restaurante compró la sopa que la gente comió.  
   *The restaurant bought the soup that the people ate*

2. El departamento revisó el avión que el mecánico arregló.  
   *The department checked the plane that the mechanic fixed*

3. El cine mostró el documental que el crítico odiaba.  
   *The movie theater showed the documentary that the critic was hating*

4. La facción detestaba la maniobra que el político ejecutaba.  
   *The faction detested the manoeuvring that the politician was doing*

5. El ciudadano creyó el discurso que el político dio.  
   *The citizen believed the speech that the politician gave*

6. La abogada citó la evidencia que el defensor negó.  
   *The lawyer cited the evidence that the defense lawyer denied*

7. El colegio diseñó el código que la administración aprobó.  
   *The school made the code that the administration approved*

8. El bombero apagó el fuego que el niño prendió.  
   *The firefighter put out the fire that the kid started*

9. El profesor tomó el café que el estudiante trajo.  
   *The professor drank the coffee that the student brought*

10. El museo cerró la exposición que el gobierno financiaba.  
    *The museum closed the exhibit that the government was financing*

11. La ciudad clausuró la playa que el huracán inundó.  
    *The city closed the beach that the hurricane flooded*

12. El gato agarró el jamón que el perro comía.  
    *The cat grabbed the ham that the dog was eating*

13. La tienda vendía la pastilla que el enfermo tomó.  
    *The store sold the pill that the patient took*

14. El diario publicó la noticia que el estudiante leyó.  
    *The newspaper published the news that the student read*

15. El periódico publicó la noticia que el redactor escribió.  
    *The newspaper published the news that the editor wrote*
BIBLIOGRAPHY


Bender, Angela D., Hannah L. Filmer, Claire K. Naughtin, and Paul E. Dux (2017). Dynamic, continuous multitasking training leads to task-specific improvements but does not transfer across action selection tasks. npj Science of Learning, 14(2). doi:10.1038/s41539-017-0015-4


Coetzee, Andries W (2009). Grammar is both categorical and gradient.


Dowens, Margaret Gillon, Marta Vergara, Horacio A. Barber, and Manuel Carreiras (2010). Morphosyntactic Processing in Late Second-Language Learners. *Journal of Cognitive Neuroscience, 22*(8): 1870–1887


Lyons, Ian M., and Sian L. Beilock (2012). Mathematics Anxiety: Separating the Math from the Anxiety, Cerebral Cortex, 22(9): 2102–2110,


Ospina, Raydonal, and Silvia LP Ferrari (2010). Inflated beta distributions. Statistical Papers, 51(1); 111.


Pascalis, Vilfredode and Alessandra Morelli (1990). Event-related potentials (ERPs) to binaurally presented tone pips were elicited while male and female subjects with high and low level of anxiety were engaged in sense and nonsense word recognition tasks. *Personality and Individual Differences*, 11(7): 741-749


Pozniak, Céline, Jiaying Huang, and Barbara Hemforth. (2017). Relative clause processing, structural and linear distance matter-Evidence from Mandarin, Cantonese and English Visual World experiments.


Qu, Li, Joel Jia Wei Low, Ting Zhang, Hong Li, and Philip David Zelazo (2015). Bilingual advantage in executive control when task demands are considered. *Bilingualism: Language and Cognition, doi:10.1017/S136678914000376*


Sevilla, Yamila, Mora Maldonado, and Diego E. Shalóm.. (2014). Pupillary dynamics reveal computational cost in sentence planning. doi:10.1080/17470218.2014.911925


