THE NEUROCOGNITION OF ADULT SECOND LANGUAGE LEARNING: AN FMRI STUDY

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Linguistics

By

Kaitlyn M. Tagarelli, M.S.

Washington, DC
August 28, 2014
THE NEUROCOGNITION OF ADULT SECOND LANGUAGE LEARNING:
AN FMRI STUDY

Kaitlyn M. Tagarelli, M.S.

Thesis Advisor: Michael T. Ullman, Ph.D.

ABSTRACT

This study investigates the neural structures and cognitive processes involved in adult second language (L2) learning, and whether and how they change as a function of increasing exposure and proficiency. By combining behavioral and functional magnetic resonance imaging (fMRI) measures, this study aims to address fundamental questions in Second Language Acquisition research that cannot be fully explored with behavioral methods alone.

Research to date suggests that the mechanisms underlying L2 may change during acquisition, though in different ways for lexical/semantics and grammar. However, it is logistically impossible to longitudinally investigate the course of learning a natural language from initial exposure to advanced proficiency. This has left major gaps in this research. Studies on the learning of artificial linguistic systems (e.g., artificial grammars) have begun to address this issue, but their generalizability to natural languages has been questioned. The current study aims to bridge the gap between artificial linguistic systems and natural languages by longitudinally examining the learning of a reduced natural language, or “mini-language”.

Fifteen native speakers of English were trained on a subset of Basque, from initial exposure to high proficiency. Behavioral and fMRI measures were continuously acquired during all grammar training.

Learners achieved very high proficiency in vocabulary and reasonably high proficiency in grammar, though morphosyntactic agreement was difficult to master. FMRI activation was
found in areas associated with first language (L1) processing (e.g., BA45/47, and parietal cortex for lexical/semantics, and BA44 and 6 for grammar), suggesting that late-L2 learners have access to L1 regions. Additional areas were engaged, suggesting that L1 mechanisms are not sufficient for L2 learning and processing. At early stages of learning, hippocampal activation was found for both vocabulary and grammar. At later stages, basal ganglia activation was observed for grammar, particularly in the caudate nucleus. The findings suggest that early word and grammar learning relies on declarative memory (and more explicit processes), but that grammar later relies on procedural memory (and more implicit processes). These results highlight the utility of a mini-language model, have implications for neurocognitive theories of L2, and demonstrate the importance of integrating neural and behavioral methods in L2 research.
DEDICATION

This dissertation is dedicated to my family, who has helped me to become the person I am today, and to recognize the person I want to be.
Acknowledgements

This dissertation would not have been possible without the professional and emotional support of many people. While I’m sure this list doesn’t contain every single person who helped me in some way, there are several people that I would like to acknowledge for the impact they have had on me leading up to the writing of this dissertation, and for the great appreciation I feel toward all of them.

First and foremost, I would like to thank my dissertation committee for their continued support and guidance through this process. I would like to thank my mentor, Michael Ullman, for working tirelessly with me not only from the first stages of this project, but from my first days at Georgetown. Michael, I have learned so much from you about thinking critically and doing good science, and I have cherished your advice and friendship over the years. To Alison Mackey, thank you for constantly challenging me to step out of my comfort zone, for giving me opportunities to do so, and for not letting me lose sight of the big picture. I believe that I have become a better and more versatile researcher because of you. To Elissa Newport, it has been an honor to work with you, and I truly appreciate the time and expertise that you have dedicated to this project. Thank you so much for taking an interest in this project and allocating funds to make it happen. To Kara Morgan-Short, you were a wonderful role-model for me before we ever met, and I strive to emulate your approach to interdisciplinary research. Thank you for your careful feedback, and for always being available to help me work through issues, big and small, from experimental design to interpretation of results. And to Aron Barbey, thank you for introducing me to the world of fMRI, and for always being so supportive and encouraging. I am endlessly grateful to all five of you for your efforts to see me through this dissertation, and I admire your
passion and dedication to research. You have all helped, in one way or another, to not only make this dissertation the best that it can be, but also to motivate me to continue to grow as a scientist. For this, I thank you from the bottom of my heart.

I knew very little about the ins and outs of fMRI research before I started my dissertation, and I could not have figured it all out on my own. I would like to emphatically thank Xiong Jiang, who is responsible for most of the technical training that I received in fMRI over the past couple of years. Xiong, it is hard to imagine how I would have gotten this study off the ground and into an MRI scanner without you, or how long it would have taken to learn how to analyze data without your help. You so willingly dedicated your time to help me with the smallest details, and always with a friendly smile. I would also like to thank John VanMeter, who not only helped me build a foundational knowledge about fMRI research, but also very patiently helped me find my place at the CFMI, and made sure I was able to get the data I needed. To all of the RAs at the CFMI, thank you so very much for running my subjects with me, troubleshooting technical difficulties, and keeping me company during many hours of data collection. I really enjoyed the time I spent with you at the CFMI. Thank you to Peter Turkeltaub, Aaron Newman, and Max Riesenhuber, who were on hand to provide feedback and help me work through various experimental design issues in the earlier stages of this project.

I have had the good fortune to have had many wonderful professors throughout my graduate career who have helped me build a foundational knowledge in linguistics and cognitive science, in and out of the classroom, and foster my love of learning – Hector Campos, Rusan Chen, Jeff Connor-Linton, Robert DeKeyser, Heidi Hamilton, Donna Lardiere, Ron Leow, Michael Long, Lourdes Ortega, Paul Portner, Cristina Sanz, Mark Sicoli, Anna Marie Trester,
and Lisa Zsiga. A special thank you goes out to Patrick Rebuschat – you were my first mentor at Georgetown, and from our very first meeting, you have encouraged me to work hard and make the most of my academic career. I would also like to thank some of my professors from Union College – Rob Olberg, Cheikh Ndiaye, Michelle Chilcoat, and Charles Batson – who really instilled in me a love of research, neuroscience, and languages, and a passion to go out and learn more. Every student should be lucky enough to learn from people like you. And to all of the amazing teachers that I have had throughout my life, of which there are far too many to name – you are the people who inspired me to never stop learning and to share what I’ve learned with anyone who will listen. Thank you.

Thank you to the administrative staff in the Department of Linguistics at Georgetown, especially Manela Diez and Erin Esch. I hope you know how much you are appreciated.

I am so thankful for the community of Basque speakers that I have met through this dissertation. First and foremost, thank you to Itziar Laka, who answered so many questions about the Basque language and put me in touch with native speakers, which was immensely helpful in my stimulus design process. Itziar, it has been such a pleasure getting to know you and working with you. Thank you to Aitzi Delgado Monreal, who met with me countless times while I was designing Mini-Basque, and was always so warm and friendly and helpful, and to Miren Urteaga and Yolanda Acedo for recording stimuli for me in the Basque Country. Thank you to the members of the Washington, DC Euskal Éxtea (Basque Club), especially Igan Erostarbe and Manu Suarez, who helped me learn some Basque. Eskerrik asko!

I have had the great pleasure of working with brilliant people in the Brain and Language Lab. I thank all of you, especially Laura Babcock, Kaitlyn Litcofsky, Goldie Ann MacQuaid,
Scott Miles, and Mariel Pullman – you are all exemplary scholars and wonderful friends. I would like to especially thank Jarrett Lovelett, whose intelligence, helpfulness, and friendship was invaluable over the past three years while I was working on this dissertation. I would also like to thank the undergraduate students who have helped out in various ways – Tim Dougherty, Sari Frankle, Robin Go, Natasha Janfaza, and Karen Luo. I hope you learned something!

In my time at Georgetown, I have had the great privilege to work among fellow students who amaze and inspire me every day with their intelligence, passion, and dedication to research and learning. I have made so many wonderful friends, and I could not have made it this far without their camaraderie and support. Thanks to my dissertation support group, Jessi Grieser, Julie Lake, and Amelia Tseng, and to a few special crossovers from the “Spanish Mafia,” Jessie Cox, Sarah Grey, and Ellen Serafini. It is an honor to be in the presence of such strong, intelligent women, and you have become my best critics and strongest supporters over the years. To the members of the Georgetown Psycholinguistics Lab, especially Luke Amoroso, Phillip Hamrick, Katie Kim, Nick Pandza, Lauren Park, Rebecca Sachs, and Mari Sakai, and to so many other fellow graduate students, past and present – Carlos Balhana, Annie Calderón, Marissa Fond, Daniel Ginsberg, Kathryn Jergovich, Natalia Jacobson, Eli Kissling, Narges Mahpeykar, Charlie Nagle, Sheena Shah, Corinne Seals, Laura Siebecker, Kerstin Sondermann, Julio Torres, Laura West, Nicole Ziegler (and too many more to name!) – I am so thankful for you for so many reasons. A special thanks to Shuo Zhang for very kindly offering to write Praat scripts for me. There are two graduate students who only came into my life in the past year, but who deserve a special thank you. Janire Zalbidea and Kyle Shattuck, your expertise in Basque and fMRI, respectively, and your willingness to share your time and knowledge with me, has been
remarkably helpful in the final stages of my dissertation. Thank you so very much. You have been lifesavers and miracles.

Thank you to all of my fantastic friends outside of graduate school, who have lent emotional support and welcome distractions throughout my dissertation process, and who have understood when I wasn’t as good at staying in touch or had to stay in and work. Thank you to my amazing family – the best family anyone could ever ask for – especially Mom, Dad, Maggie, Victoria, and Joey. Thank you for understanding that I needed to do this, for giving me so much to look forward to during breaks from work, and for not letting me lose sight of the important things in life. I love you all so much. Thank you to my wonderful in-laws, Judy, Dale, and Sean. I am so fortunate to have become part of your family, and relaxing and laughing with you all over the years has provided such a wonderful refuge from work. And of course, thank you to my wonderful husband, Kyle. Thank you for cooking so many delicious meals and taking on a disproportionate amount of the household chores so that I could get work done or just relax. Thank you for understanding when I had to spend so many evenings and weekends working, and for giving me a reason to take breaks and making those breaks so enjoyable. Thank you for listening whenever I needed to ramble on about something really “interesting” about linguistics, or just work through a problem out loud. You have been my sounding board, my editor-in-chief, my shoulder to cry on, my biggest fan, my shining star, and my best friend. And, going above and beyond normal spousal duties, you patiently lent your artistic abilities to draw and revise pictures of the animals for the Mini-Basque stimuli, and to really clean up Figure 11 for me. I love you so much, and I could not have done this without you standing beside me.
Finally, this dissertation would not have been possible without the support of people and organizations who thought it was a worthwhile project. Thank you to the William Orr Dingwall Foundation, who supported my final year of graduate work with a Neurolinguistics Dissertation Fellowship, the Language Learning journal, Modern Language Journal, and Cosmos Club Foundation for dissertation grants, Elissa Newport with the Center for Brain Plasticity and Recovery for providing funding for fMRI scanning, and John VanMeter with the Center for Functional and Molecular Imaging for helping me to make the most of the funding I had.

And last but not least, I would like to thank all of the Mini-Basque learners who dedicated hours of their time to participating in this study. You are amazing!
# TABLE OF CONTENTS

## CHAPTER 1: INTRODUCTION & REVIEW OF LITERATURE .......................................................... 1
  1.1 Introduction ..................................................................................................................... 1
  1.1.2 The Problem .............................................................................................................. 2
  1.2 Review of Literature ...................................................................................................... 4
    1.2.1 Theories of L2 Neurocognition .............................................................................. 4
    1.2.2 Behavioral Studies of L2 Acquisition and Processing ............................................ 9
      1.2.2.1 Artificial linguistic systems ............................................................................. 14
        1.2.2.1.1 Word learning .......................................................................................... 15
        1.2.2.1.2 Word segmentation ................................................................................ 16
        1.2.2.1.3 Artificial grammar learning .................................................................. 17
        1.2.2.1.4 Artificial language learning .................................................................. 18
        1.2.2.1.5 Mini-language learning ......................................................................... 19
      1.2.3 Electrophysiological Evidence: Event-Related Potentials (ERPs) ..................... 21
        1.2.3.1 Artificial linguistic systems ........................................................................... 26
          1.2.3.1.1 Word learning ........................................................................................ 26
          1.2.3.1.2 Word segmentation .............................................................................. 26
          1.2.3.1.3 Artificial grammar learning ................................................................ 27
          1.2.3.1.4 Artificial language learning ................................................................ 27
          1.2.3.1.5 Mini-language learning ....................................................................... 29
      1.2.4 Functional Neuroanatomy: PET and fMRI ......................................................... 32
        1.2.4.1 Artificial linguistic systems .......................................................................... 34
          1.2.4.1.1 Word learning ....................................................................................... 34
          1.2.4.1.2 Word segmentation .............................................................................. 35
          1.2.4.1.3 Artificial grammar learning ................................................................ 35
          1.2.4.1.4 Artificial language learning ................................................................ 37
          1.2.4.1.5 Mini-language learning ....................................................................... 37
      1.2.5 Summary of Evidence ............................................................................................. 38
  1.3 The Current Study ......................................................................................................... 39
    1.3.1 Research questions ................................................................................................. 40
    1.3.2 Hypotheses ............................................................................................................. 40

## CHAPTER 2: RESEARCH METHODS AND DESIGN ................................................................. 44
2.1 Participants .................................................................................................................. 44
  2.1.1 Language background ....................................................................................... 45
2.2 Materials ................................................................................................................... 45
  2.2.1 The mini-language: Mini-Basque ....................................................................... 45
    2.2.1.1 The source language: Basque .................................................................. 46
      2.2.1.1.2 Choosing a source language ............................................................... 46
      2.2.1.1.3 Features of Basque .......................................................................... 48
    2.2.1.2 Designing Mini-Basque ........................................................................... 50
      2.2.1.3 Features of Mini-Basque ...................................................................... 52
  2.2.2 Auditory stimuli ................................................................................................... 56
  2.2.3 Tasks ..................................................................................................................... 58
    2.2.3.1 Picture matching ......................................................................................... 59
    2.2.3.2 Control (tone-matching) task .................................................................... 65
  2.2.3.3 Timed grammaticality judgment task ............................................................ 67
  2.2.3.4 Production ...................................................................................................... 71
    2.2.3.5 Additional assessments ............................................................................. 72
      2.2.3.5.1 Written grammaticality judgment task ............................................... 72
      2.2.3.5.2 Free response task .............................................................................. 73
      2.2.3.5.3 Debriefing questionnaire .................................................................... 74
      2.2.3.5.4 Wrap-up questionnaire ....................................................................... 74
      2.2.3.6 Demographic information ...................................................................... 75
  2.3 Pilot testing ............................................................................................................... 75
    2.3.1 Behavioral Pilot 1 ............................................................................................ 75
    2.3.2 Behavioral Pilot 2 ............................................................................................ 76
    2.3.3 Behavioral Pilot 3 ............................................................................................ 78
    2.3.4 Behavioral Pilot 4 ............................................................................................ 80
  2.4 Procedure .................................................................................................................. 81
    2.4.1 Day Ø: Screening and cognitive testing ............................................................ 83
    2.4.2 Day 1: Training session 1 ................................................................................. 83
    2.4.3 Day 2: Training session 2 ............................................................................... 84
    2.4.4 Day 3: Training session 3 ............................................................................... 84
    2.4.5 Retention ......................................................................................................... 84
  2.5 Behavioral data analysis ........................................................................................... 85
4.4 Future Directions and Conclusions ................................................................. 168
APPENDIX A ........................................................................................................ 175
APPENDIX B ........................................................................................................ 176
APPENDIX C ........................................................................................................ 178
APPENDIX E ........................................................................................................ 188
APPENDIX F ........................................................................................................ 190
REFERENCES ...................................................................................................... 191
LIST OF TABLES

Table 1. The realizations of the Basque determiner according to case and number....................... 55
Table 2. The realizations of the Basque auxiliary, *ukan*, according to polypersonal agreement.. 55
Table 3. Accuracy on the word-level picture-matching task ..................................................... 97
Table 4. Reaction times on the word-level picture-matching task............................................. 100
Table 5. Accuracy on the sentence-level picture-matching task ................................................. 105
Table 6. Reaction time on the sentence-level picture-matching task ......................................... 107
Table 7. Accuracy and *D’* scores for the grammaticality judgment task................................. 111
Table 8. Reaction time for grammaticality judgment task......................................................... 116
Table 9. Performance on the word production task................................................................... 118
Table 10. Performance on the sentence production task............................................................. 119
Table 11. Peaks for contrasts between the sentence learning task and baseline....................... 121
Table 12. Peaks for contrasts between the sentence learning task and baseline on Day 1. ....... 123
Table 13. Peaks for contrasts between the sentence learning task and baseline on Day 3. ....... 125
Table 14. Peaks for contrasts between the lexical condition and baseline over all runs. ........... 126
Table 15. Peaks for contrasts between the agreement condition and baseline over all runs ...... 127
Table 16. Peaks for contrasts between the word-order condition and baseline over all runs..... 128
Table 17. ROI analyses for the sentence-level picture-matching task........................................ 132
Table 18. Significant and trending correlations between accuracy and activation across runs. . 137
Table 19. Correlations between delta *d’* values and activation for syntactic and morphosyntactic violations................................................................................................. 139
List of Figures

Figure 1. The Basque Country of Europe ................................................................. 49

Figure 2. Sample images and citation forms for the 12 open-class content words and corresponding images in Mini-Basque. ................................................................. 53

Figure 3. Sample Mini-Basque sentence with associated images ................................ 56

Figure 4. Praat script arrangement of pauses and phrases in Mini-Basque sentences ........ 58

Figure 5. Timing of word-level picture-matching task ............................................. 61

Figure 6. Timing of sentence-level picture-matching task ....................................... 63

Figure 7. Timing of control task .............................................................................. 67

Figure 8. Timing of timed grammaticality judgment task ...................................... 70

Figure 9. Word and sentence production ............................................................... 72

Figure 10. Free response task .................................................................................. 74

Figure 11. Detailed fMRI procedure ....................................................................... 82

Figure 12. Accuracy on the word-level picture matching task ............................... 96

Figure 13. Response time on the word-level picture matching task ....................... 99

Figure 14. Accuracy on the word-level picture matching task over 40 mini-blocks .... 102

Figure 15. Response time on the word-level picture matching task over 40 mini-blocks .... 103

Figure 16. Accuracy on the sentence-level picture matching task .......................... 104

Figure 17. Response time on the sentence-level picture matching task .................. 106

Figure 18. Accuracy on the sentence-level picture matching task over 32 mini-blocks .... 109

Figure 19. Response time on the sentence-level picture matching task over 32 mini-blocks .... 110

Figure 20. $D'$ values for the grammaticality judgment task ........................................... 112
Figure 21. Response time on the grammaticality judgment task ........................................ 117
Figure 22. Performance on the word production task ..................................................... 118
Figure 23. Performance on the sentence production task .............................................. 119
Figure 24. Sentence-level activation in the whole-brain analysis .................................. 124
Figure 25. Word-level activation in ROI analyses ......................................................... 130
Figure 26. Changes in activation in (a) L BA44 and (b) L BA6 at the word level .......... 130
Figure 27. Activation for syntactic violations versus grammatical sentences ............... 134
Figure 28. Changes in activation in the right putamen for syntactic violations ............. 134
Figure 29. Scatterplots of the beta values against the arcsine of accuracy for the first word-level picture-matching run ................................................................. 136
Figure 30. Activation-likelihood estimation meta-analysis of L1 and L2 grammar activation .. 148
Chapter 1: Introduction & Review of Literature

1.1 Introduction

This study investigates the neurocognitive mechanisms involved over the course of learning a second language (L2) in adulthood. Throughout this dissertation, I aim to address several questions within this area, most broadly: How does L2 acquisition actually proceed in the mind and brain? More specifically, which brain structures and associated computational mechanisms are involved in the acquisition of adult-learned L2 from the very beginning to high proficiency? Moreover, do the structures and cognitive mechanisms involved differ for different aspects of language, particularly lexical/semantics, syntax, and morphosyntax? Answering these questions can provide a deeper understanding of fundamental issues in Second Language Acquisition (SLA) research that cannot be answered with behavioral methods alone. This project attempts to address these questions by examining the acquisition of a simplified natural language (a mini-language) during the entire trajectory of learning, using functional magnetic resonance imaging (fMRI) and behavioral measures.

Additionally, the proposed methodology should shed light on other important questions relating to SLA and the neurocognition of L2. For example, revealing the neural regions involved in L2 learning and processing can elucidate the extent to which L2 relies on structures and mechanisms that are also available in L1 processing. Additionally, previous neuroscience research on the functions of these brain regions may provide information about the types of processes involved in L2 learning and processing, including whether these processes are implicit.

1 Throughout this dissertation, I use the terms learning and acquisition interchangeably, which is common practice in the field of cognitive neuroscience. I do not mean to imply that acquisition is spontaneous and implicit, learning is intentional and explicit, and the outcomes of each process are stored separately, as traditionally defined in SLA (Krashen, 1981).
and/or explicit. Moreover, collecting both neural and behavioral measures in a longitudinal
design will allow us to understand how these measures relate to each other over time, and for
different aspects of language. Doing so with continuous neural measures provides a more fine-
grained approach than has been employed in previous research. For example, if there is a shift in
activity from one neural structure to another as learning progresses, this design can examine,
with more precision than previous research, when this shift occurs and whether it differs
according to language domain. Finally, this paradigm provides a useful framework from which
to examine other questions in L2 acquisition research, such as the role of training conditions
(e.g., immersion-like vs. classroom-like) and individual differences in cognitive abilities in
predicting language learning processes and outcomes.

1.1.2 The Problem. Adult L2 acquisition, as opposed to L1 acquisition or child L2
acquisition is notoriously difficult. This is a commonly accepted fact among the general
population and a finding that is largely agreed upon among scholars in the field of L2
acquisition, though the reasons for the negative relationship between the efficiency of learning
and age of acquisition remain contested (Abrahamsson & Hyltenstam, 2009; Babcock, Stowe,
Maloof, Brovetto, & Ullman, 2012; Birdsong, 2006; DeKeyser, 2000; Ioup, Boustagui, El Tigi,
& Moselle, 1994; Johnson & Newport, 1989; Lenneberg, 1967; Marinova-Todd, Marshall, &
Snow, 2000; Mueller, 2006). This dissertation focuses specifically on late-L2 learners, which are
not to be confused with, and may be fundamentally different from, early-L2 learners. In a similar
vein, early-L1 learners should not be assumed to rely on the same mechanisms as late-L1
learners (e.g., deaf individuals who learn sign language past adolescence), and there is indeed
evidence that highlights the fallacy of such an assumption (Mayberry, Chen, Witcher, & Klein,
In sum, there is not a consistent one-to-one relationship between early language learning and L1, and late language learning and L2. L2s can be learned early, L1s can be learned late, and the mechanisms involved in learning any language may have to do with both age effects and the order in which the languages are acquired, which are distinct variables. Throughout this dissertation, unless otherwise noted L1 refers to early-learned L1, and L2 refers to late-learned L2, acquired past adolescence. Adult L1 processing, unless otherwise noted, refers to processing of an early-learned L1 that occurs later in life, i.e., in adulthood.

While research in SLA and cognitive neuroscience has come a long way in characterizing the process of adult L2 acquisition, methodological limitations have left gaps in the literature. The biggest roadblock to fully understanding L2 acquisition is the amount of time that it takes to learn an L2. People spend years and years learning second languages, and often never reach native-like proficiency (e.g., Abrahamsson & Hyltenstam, 2009; DeKeyser, 2000; Johnson & Newport, 1989). It is logistically impractical to follow learners for years over the course of their L2 learning process, and it is therefore rarely done (but see, for example, Lardiere, 1998). Therefore, research into natural L2 acquisition can tell us about processing at snapshots, like low and high proficiency (Bowden, Steinhauer, Sanz, & Ullman, 2013), or how processing changes in a brief period of time, like one semester (J. McLaughlin, Osterhout, & Kim, 2004), but it necessarily misses important aspects of the full trajectory of learning.

To address some of these limitations, researchers have turned to artificial linguistic systems, such as learning of word lists, word segmentation, artificial grammar learning, and artificial language learning. These are models of language that can be learned in a short period of time, usually over the course of several hours. However, because they are artificial, the extent to
which findings from these studies can generalize to natural language learning is unclear. By
investigating longitudinal and continuous learning of a mini-language, which contains real words
and structures that exist in a natural language, this study aims to address gaps and add to the
literature on the course of L2 acquisition in the mind and brain.

1.2 Review of Literature

As mentioned above, much of what we do know about language learning and processing
comes from studies of natural language learning, but methodological limitations associated with
studying natural languages have made it difficult to answer certain questions about L2
acquisition. For this reason, many researchers in psychology and SLA have turned to artificial
linguistic systems, including the learning of word lists, word segmentation from a speech stream,
artificial grammars, artificial languages, and mini-languages. Moreover, the use of a variety of
methods, including behavioral, electrophysiological, and neuroanatomical measurements, as well
as cross-sectional and longitudinal designs, has broadened our understanding of adult L2
acquisition. Nevertheless, significant gaps remain.

This section provides an overview of three prominent groups of neurocognitive theories
of L2 learning and processing, and their predictions, which draw on current evidence from L2
research. It then reviews the current literature in L2 acquisition and processing for both natural
and artificial linguistic systems, looking at behavioral, event-related potential (ERP), and
functional neuroanatomical studies, including cross-sectional, longitudinal, and continuous
designs.

1.2.1 Theories of L2 Neurocognition. A growing body of literature has investigated the
neurocognitive processes involved in L2 acquisition (for reviews, see (Abutalebi, 2008; Clahsen,
Felser, Neubauer, Sato, & Silva, 2010; Doughty & Long, 2005; Gor, 2010; Green, 2003; Hernandez & Li, 2007; Indefrey, 2006; Kotz, 2009; Kroll & de Groot, 2005; Morgan-Short & Ullman, 2012; Paradis, 2009; Schmidt & Roberts, 2009; Steinhauer, White, & Drury, 2009; Ullman, 2005). Two important questions that remain regarding the neural and computational aspects of L2, which are of particular interest here, include whether L2 and L1 depend on the same or different neurocognitive mechanisms, and whether the degree of overlap, if any, between L1 and L2 processes vary depending on the language functions involved (e.g., lexical/semantics, syntax, and morphosyntax). There are three broad theoretical frameworks that attempt to characterize the neurocognition of L2 learning and processing.

First, it has been claimed that L2 and L1 rely on fundamentally distinct neural substrates (Albert & Obler, 1978; Bley-Vroman, 1989, 2009; Hagen, 2008). According to this view, children have access to innate and efficient language learning mechanisms when learning their L1, but these mechanisms are no longer available after a certain critical period. The L2 must therefore be learned using different mechanisms, and thus different neural substrates. There has been little support for the strong view of this theory, known as the Fundamental Difference Hypothesis (Bley-Vroman, 1989), and this hypothesis has been recently revised to suggest that L1 processing primarily relies on language specific processes, whereas L2 processing relies on a variety of mechanisms, all of which are recruited, to a lesser extent, by native language processes (Bley-Vroman, 2009). This view is rather non-specific in terms of neural predictions, with most proponents simply positing a dissociation between brain structures involved in L1 and L2 processing (Albert & Obler, 1978; Hagen, 2008).
Other researchers have argued that L2 and L1 depend on largely overlapping structures and mechanisms (Abutalebi, 2008; N. C. Ellis, 2005; Green, 2003; Indefrey, 2006; Hernandez, Li, & MacWhinney, 2005; MacWhinney, 2011; Perani & Abutalebi, 2005). These theories claim that L2 is acquired by a neural system that has already been specified for learning L1. One of the most well-developed models in this group is the Convergence Hypothesis (Abutalebi, 2008; Green, 2003; Perani & Abutalebi, 2005). According to this hypothesis, both lexical/semantics and grammar in L1 and L2 should activate areas in the left inferior frontal gyrus (LIFG), including Brodmann Areas (BA) 44, 45, and 47. Temporo-parietal regions are also recruited for lexical/semantics in L1 and L2.

Nonetheless, some differences are expected between L1 and L2 activation, particularly at low proficiency levels. In particular, more extensive activation in the LIFG, as well as additional activation in the prefrontal cortex and basal ganglia, is expected in L2 processing as compared to L1 processing. In this model, this additional activation is not associated with the L2 language network, but rather represents processing-specific demands associated with cognitive effort. It has been hypothesized that the differences between L1 and L2 activation are due to lower efficiency of L2 processing (Indefrey, 2006), lower efficiency of neural organization in L2 (Indefrey, 2006), or the greater role of executive control in a “weaker” L2 system (see Abutalebi, 2008). The Convergence Hypothesis predicts that as L2 learners become more proficient, L2 activation patterns converge with L1 activation patterns. That is, activation in the PFC and basal ganglia should decrease, and activation in the LIFG should be more concentrated around L1 areas. This decrease in activation reflects a change from controlled to automatic processing. While proficiency is cited as the key factor that determines the convergence of L1 and L2
lexical/semantics, this hypothesis suggests that age of acquisition (AoA) is a more important factor in determining the neural correlates of L2 grammar processing. That is, the later an individual learns an L2 (particularly after some critical period), the less convergent the L1 and L2 grammar activation will be. However, AoA and proficiency are highly confounded variables, as individuals who learn languages early in life tend to reach higher proficiency than later learners. For this reason, proponents of this hypothesis do not make strong claims about the role of either AoA or proficiency in the neural bases of L2 grammar. In any case, in the present study, all learners are young adults, so effects due to AoA are not anticipated.

A third and more intermediate view posits a reliance on substrates different from those involved in L1 in the initial stages of L2 acquisition, and a shift to reliance on L1 neurocognition as a result of increasing exposure or proficiency in L2 (Clahsen & Felser, 2006b; Paradis, 2009; Ullman, 2001c). Theories within this third view differ with respect to their predictions for lexical and grammatical functions. Whereas Paradis (1994, 2004, 2009) posits a comparable shift in reliance for both lexical and grammatical functions, Clahsen and Ullman suggest that lexical/semantics relies on similar structures in L1 and all stages of L2, and the shift in reliance from L2 to L1 substrates only occurs for grammar (Clahsen & Felser, 2006a, 2006b; Ullman, 2001b, 2001c, 2004, 2005, 2006, 2012). Of these three theories, Ullman’s (Ullman, 2001b, 2001c, 2004, 2005, 2006, 2012) Declarative/Procedural (DP) Model has most clearly specified the underlying functional neuroanatomy of these patterns.

According to the DP model, language learning and use depend on two long-term memory brain systems, declarative and procedural memory. Lexical/semantics should always depend on declarative memory, which seems to be specialized for learning idiosyncratic knowledge, a core
component of lexical/semantics. Learning in declarative memory initially depends strongly on
the hippocampus and other medial temporal lobe (MTL) structures. Knowledge and its
processing in declarative memory eventually rely more on temporal and other neocortical
regions, including inferior frontal regions for recall (e.g., BA45/47; Eichenbaum, 2012; Ullman,
2004; Wixted & Squire, 2011), and the posterior parietal lobe. This functional neuroanatomy is
predicted for lexical/semantics in both L1 and L2.

Grammar, in contrast, should be learnable by both declarative and procedural memory
(Ullman, 2004, 2005). The procedural memory system is rooted in frontal/basal ganglia circuits
and seems to be specialized for learning statistical regularities, which are gradually acquired with
repeated exposure. The basal ganglia appear to be involved primarily in skill acquisition, while
(pre)motor and related regions, including BA44 and BA6, may be more important for processing
automatized skills (Ashby, Turner, & Horvitz, 2010; Doyon et al., 2009). Declarative memory,
however, is highly flexible and can also learn aspects of grammar. The DP model predicts that in
L2, grammatical abilities will initially depend largely on declarative memory, mainly because
this system learns faster than procedural memory. However, procedural memory should
gradually learn the grammar, and should eventually take precedence as a more efficient system
for sequential processing (Ullman, 2004). Therefore, L2 grammar learning at low proficiency
should involve both declarative memory brain structures and the basal ganglia, with gradually
increasing engagement of BA44 as exposure and proficiency increase (Ullman, 2004). In
contrast, like adult L1 speakers, very high proficiency L2 speakers are expected to rely mainly
on BA44 and related premotor regions (BA6) for processing automatized grammar.

These three models, the Fundamental Difference Hypothesis, Convergence Hypothesis,
and Declarative/Procedural Model, make notably different claims about the nature of L2 learning and processing, particularly with respect to the neural overlap of L1 and L2, low and high proficiency, and lexical/semantics and grammar. All three theories focus on late-learned L2s, but while age is an important factor in all three groups of theories, proficiency is also expected to play a role in both the Convergence Hypothesis and the DP Model, where individuals with high proficiency may overcome some of the effects of age in L2 processing. Importantly, these three perspectives make distinct predictions about the nature of L2 learning and processing, and can therefore be empirically distinguished.

The following section provides a review of the findings in empirical L2 acquisition research, which has come a long way in informing the development of the abovementioned theories. I discuss findings from behavioral, electrophysiological, and functional neuroanatomical research, focusing first on studies of natural L2 learning and processing and then artificial linguistic systems and mini-languages for each technique.

1.2.2 Behavioral Studies of L2 Acquisition and Processing. While the field of SLA is a relatively young one, without a unifying theory of how L2 learning occurs, there has been much progress in the past 40 years or so in developing an understanding of this process. It is clear that unlike L1 acquisition, L2 acquisition usually takes a considerable amount of effort, and is not always successful. VanPatten & Williams (2007) provide a list of several observations that have been made about L2 acquisition as a result of the empirical research to date. Among these are the observations that there is variability in the extent to which different linguistic subsystems (e.g., lexical/semantics, phonotactics, syntax) are developed at any given time, that there are
predictable developmental stages that learners go through when acquiring certain structures, and that the success of L2 acquisition is highly variable across learners.

While adult L2 acquisition is a generally difficult task, some aspects of language seem to be easier to learn than others. In particular, lexical/semantic aspects are easier to acquire than syntactic and phonological aspects of L2 (see de Groot & Kroll, 1997; Ritchie & Bhatia, 1996). Yet even within the domain of (morpho)syntax, some features of language are more difficult to acquire than others (DeKeyser, 2005; Johnson & Newport, 1989; Newport, 1990). In a comprehensive review of the factors that determine the difficulty of acquiring grammatical forms, DeKeyser (2005) claims that basic word order is acquired relatively quickly with few problems after the initial stages of learning, but that morphosyntax is very difficult for L2 learners, especially when it is complex, as in richly inflected languages. Learners tend to demonstrate problems in L2 morphology (Jiang, 2004; MacWhinney, 2001; VanPatten, 2004; Yeni-Komshian, Robbins, & Flege, 2001) even after many years of L2 exposure (DeKeyser, 2000; Johnson & Newport, 1989). This pattern of successful learning of word order but difficulty with morphology has also been demonstrated in late-L1 learners who acquired American Sign Language in adulthood (Newport, 1990), suggesting that it may have more to do with age effects than with learning a subsequent language.

Particularly within the domain of morphosyntax, there is substantial evidence for predictable developmental sequences for L2 learners. Seminal work by Brown (1973) demonstrated that children acquire morphemes in a particular order when learning their L1. For example, present progressives and the irregular past tense are learned relatively early, while articles and the regular past tense are learned later, and third person irregular inflection and
contractions are learned later still. Extending this work to L2, morpheme order studies have established an order of acquisition of morphemes in L2 English (Andersen, 1977; Bailey, Madden, & Krashen, 1974; Dulay & Burt, 1973, 1974; Fathman, 1975; Gass & Selinker, 2001; Goldschneider & DeKeyser, 2001; Houck et al., 1978; Krashen, Sferlazza, Feldman, & Fathman, 1976; Larsen-Freeman, 1975, 1976; Larsen-Freeman & Long, 1991; Long & Sato, 1984; Pica, 1983; Rosansky, 1976), and research on the acquisition of Hebrew morphology suggests that salience is a key factor in determining which morphemes will be acquired first (DeKeyser, Alfi-Shabtay, & Ravid, 2010). Following Brown’s (1973) work on L1 morphology acquisition and drawing on research into L2 morpheme acquisition order, DeKeyser (2005) argues that difficulty of L2 morphology acquisition depends on the complexity of the form, the complexity of the meaning, and the complexity of the form-meaning relationship to be learned. He cites evidence that certain forms, including articles, classifiers, grammatical gender, and verbal aspect, are particularly difficult for speakers to acquire because “they express highly abstract notions that are extremely hard to infer, implicitly or explicitly, from the input” (p. 5). These forms are even more difficult if they are novel, that is, not in the learner’s L1 background, which suggests a role for L1 transfer effects. Adult learners, in particular, seem to have a great deal of difficulty learning grammatical dependencies, including agreement (e.g., subject-verb, adjective-noun), noun declensions, and verb conjugations (Braine, 1987; Braine et al., 1990; Brooks, Kempe, & Sionov, 2006; Newport & Aslin, 2004).

Finally, DeKeyser (2005) argues that morphological features that are redundant, optional, or opaque are especially difficult to learn. This idea is in line with Input Processing (VanPatten, 1990, 2007), which suggests that a morphological form may not be necessary for understanding
the meaning of a sentence when its meaning is expressed by another element in the sentence. For example, when salient temporal adverbs such as “yesterday” or “tomorrow” are present in a sentence, less salient verb tense markers are redundant and therefore more difficult to learn. Optional constructions, such as null subjects in Spanish or Italian, suggest redundancy and are therefore difficult to learn for many of the same reasons that redundant forms are. Forms that are opaque, such as the “–s” affix in English, may also be difficult to learn because, in this case, it maps to three different meanings (i.e., plural of the noun, 3rd-person singular verb, and genitive of the noun), each of which is realized with the same three allomorphs (i.e., [z], [s], and [iz]). Within L1, there has been some debate as to whether these same characteristics (redundant, optional, and opaque) make morphological features more difficult for children to learn (for an overview of different perspectives in this debate, see Pinker, 1987). However, there is compelling evidence to suggest that children do not, in fact, have a harder time learning these somewhat abstract morphological features, like grammatical gender (Maratsos & Chalkley, 1980), which supports the notion that children learning an L1 rely on different mechanisms and cues than adults learning an L2.

Behavioral studies of grammatical processing in L2 learners can shed light on the neurocognitive theories outlined in the previous section. In particular, the DP model (Ullman, 2001c, 2005, 2006, 2012) argues that in the L1, lexical memory is subserved by the declarative memory brain system, while complex grammatical processing depends on rule-governed computational mechanisms subserved by the procedural memory brain system. At low proficiency L2, both lexical and grammatical functions are thought to depend on the declarative memory system. Therefore, low proficiency L2 speakers should store lexical items like “walk”
and “cat,” but may also store more complex forms like “walked” and “cats” as chunks. In L1 speakers, and in higher proficiency L2 speakers, such forms are posited to be composed in the procedural memory system (i.e., “walk” + “-ed” and “cat” + “-s”). This phenomenon can be tested by frequency effects analyses, which assumes that higher frequency chunks should be processed faster than lower frequency chunks, but processing time for composed forms should not vary according to frequency.

Evidence from empirical studies investigating the processing of irregular and inflected forms suggest that irregular forms are stored in both L1 and L2 (Beck, 1997; Birdsong & Flege, 2001; Bowden, Gelfand, Sanz, & Ullman, 2010; Lalleman, Van Santen, & Van Heuven, 1997; Neubauer & Clahsen, 2009; Silva, 2009), and that regulars are stored at low proficiency L2 (Bowden et al., 2010; Neubauer & Clahsen, 2009; Silva, 2009; Silva & Clahsen, 2008) but composed at high proficiency L2 (Birdsong & Flege, 2001; Lalleman et al., 1997). A recent study (Babcock et al., 2012) specifically investigated length of residence as a variable in storage of irregular and inflected forms. Overall, L2 speakers showed evidence of storing both irregular and regular forms, while L1 speakers seemed to store irregular forms and compose regular forms. However, longer length of residence for females and later age of arrival for all learners was associated with a greater tendency to compose regulars; such effects were not seen for irregulars. These findings, along with the other empirical evidence to date, support the predictions of Ullman and Clahsen, suggesting that lexical functions in the L2 depend on the same neural underpinnings as those in L1, whereas grammatical functions first depend on distinct structures at low proficiency, but shift reliance to the L1 mechanisms as proficiency increases.
In sum, studies on the acquisition of natural languages in adulthood reveal a distinction between the acquisition and processing of lexical/semantic and grammatical aspects of the L2, and further distinctions within the realm of (morpho)syntax. Research to date on the acquisition of full natural languages provides extremely important insights about L2 processing at different proficiency levels and a sturdy foundation from which further research can be launched. However, there are some limitations associated with the study of natural languages. In particular, mastery of an L2 takes a substantial amount of time, which makes it logistically impractical to study the entire course of learning a natural language from no knowledge to high proficiency. Additionally, while it is important to get a complete picture of L2 acquisition, which necessarily involves the interaction of multiple language domains and types of processing, natural languages are not well-suited for testing more specific questions about the acquisition of lexical/semantics versus grammar, or statistical versus rule-based versus analogical learning. For this reason, some research in psychology and SLA has turned to artificial linguistic systems, including word lists, word segmentation, artificial grammars, artificial languages and mini-languages. As will be demonstrated throughout this review of literature, the study of artificial linguistic systems has corroborated much of the findings of natural language learning and broadened our understanding of adult language learning.

1.2.2.1 Artificial linguistic systems. Artificial linguistic systems are small-scale systems that can be used to test certain aspects of language in controlled laboratory settings. They constitute models for language learning that are particularly useful because they can be learned in short periods of time, and can reliably isolate different aspects of language. Importantly, certain artificial systems, like artificial languages and mini-languages, combine aspects of
language and may therefore be more generalizable to natural language learning. Below I discuss findings from five types of artificial linguistic system paradigms: word learning, word segmentation, artificial grammar learning, artificial language learning, and mini-language learning. The first three systems model domains of language in isolation, and the latter two combine various language domains.

1.2.2.1.1 Word learning. Vocabulary acquisition is a crucial part of both first and second language acquisition, and its relative simplicity compared to (morpho)syntactic acquisition has made it a frequent target of laboratory research. In addition to knowledge about phonological form, word learning crucially involves semantic information (Dell & O’Seaghdha, 1992; Posner & Carr, 1992). Importantly, word learning involves more than word-object associations; rather, it consists of learning referents for abstract concepts (Waxman & Gelman, 2009).

Children and adults have been shown to acquire words for new objects after very few presentations, a phenomenon known as “fast-mapping” (Carey & Bartlett, 1978; Heibeck & Markman, 1987; Waxman & Gelman, 2009), and to retain this knowledge over time (Markson & Bloom, 1997). A number of laboratory studies have demonstrated that learners can correctly recognize or produce words at reasonably high accuracy after minutes to hours of training (Batterink & Neville, 2011; Breitenstein et al., 2005; Hamrick & Rebuschat, 2014; Jeong et al., 2010; Mestres-Missé, Càmara, Rodríguez-Fornells, Rotte, & Münte, 2008; Mestres-Missé, Rodríguez-Fornells, & Münte, 2010; Raboyeau et al., 2004; Raboyeau, Marcotte, Adrover-Roig, & Ansaldo, 2010; Sakai, Miura, Narafu, & Muraishi, 2004; Sandak et al., 2004), and even retain some knowledge after months of no exposure (Raboyeau et al., 2004). These studies demonstrate
that word learning in adults seems to follow very similar patterns as word learning in child L1 acquisition.

1.2.2.1.2 Word segmentation. Laboratory studies in which learners are trained on word lists may bear some resemblance to vocabulary drills in L2 classrooms, but natural language word learning also crucially involves the ability to extract lexical items from the continuous stream of language input. A solid line of research has demonstrated that both children and adults are able to identify word boundaries in a continuous speech stream based on statistical information, such as transitional probabilities between syllables (Aslin, Saffran, & Newport, 1998; Cunillera et al., 2009; Karuza et al., 2013; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), though this knowledge is not necessarily explicit (McNealy, 2006; McNealy, Mazziotta, & Dapretto, 2011; Scott-Van Zeeland et al., 2010). A common paradigm in these studies involves presenting learners with a continuous speech stream made up of randomly placed trisyllabic words (e.g., “pabiku,” “golatu,” “daropi,” “tibodu”). Transitional probabilities of syllables within a word are high, whereas transitional probabilities across word boundaries are lower (e.g., 0.33), because of the random order of the words. For example, using the words above, “pa” is followed by “bi” 100% of the time, but “ku” can be followed by either “go,” “da,” or “ti.” After a short period of listening to the speech stream (often two minutes), learners demonstrate an ability to distinguish words from nonwords or partwords. Additional cues, such as prosody, have been shown to enhance the learning effect in word segmentation (Saffran, Newport, et al., 1996). Research on word segmentation has made an important contribution to the study of language acquisition by highlighting the role of statistical cues in this process. However, these studies focus on the
sequential order of linguistic elements in a sequence, which ignores the complexity of the hierarchical structure of natural languages.

1.2.2.1.3 Artificial grammar learning. Artificial grammars (Reber, 1967) are typically structured systems (e.g., letter strings, a series of tones, syllables, or pseudowords), whose order is determined by rules that may be consistent with those found in natural languages. The main advantage of studying artificial grammars over natural languages is the ability to control for previous input and the amount and type of exposure during training. Another advantage of artificial grammars is that learners can be trained in a very short period of time (i.e., minutes; Robinson, 2010). Behaviorally, studies have demonstrated robust learning effects as a result of artificial grammar training in different learning conditions, such as implicit and explicit, and for different types of artificial grammars, such as regular or finite-state systems and context-free or context-dependent grammars (Folia et al., 2008; Forkstam, Elwér, Ingvar, & Petersson, 2008; Petersson, Folia, & Hagoort, 2012; Petersson, Forkstam, & Ingvar, 2004; Reber, 1990, 1993, 2003; Reber, Kassin, Lewis, & Cantor, 1980; Reber, Walkenfeld, & Hernstadt, 1991; Robinson, 2005). There is a debate as to whether artificial grammar learning involves the implicit acquisition of complex sequential information (Reber et al., 1991) or the explicit acquisition of cues such as chunk (e.g., bigram or trigram) strength (Perruchet & Pacteau, 1990).

Linguistic artificial grammars offer an improvement over nonlinguistic artificial grammars (Fitch & Hauser, 2004; Friederici, 2004; Friederici, Steinhauer, & Pfeifer, 2002; Opitz & Friederici, 2003), and taken together, both shed light on how humans acquire and process structured sequences, which has important implications for natural language processing. Unlike natural languages, however, artificial grammars do not contain semantic information, so no
form-meaning connections exist in the grammars. As a result, they cannot be spoken or understood. Though research in artificial grammar learning is useful in the study of L2 acquisition, more natural artificial systems may provide further insight into this area.

1.2.2.1.4 Artificial language learning. Artificial and semi-artificial languages, as opposed to grammars, crucially include form-meaning relationships. Like the systems described above, they allow researchers to control for previous input and to isolate aspects of language to investigate. Semi-artificial languages (Alanen, 1995; Boyd, Gottschalk, & Goldberg, 2009; Grey, Williams, & Rebuschat, 2014; Hama & Leow, 2010; Hamrick, 2014a, 2014b; Kim & Rebuschat, in prep; Leung & Williams, 2012; Rebuschat, Hamrick, Sachs, Riestenberg, & Ziegler, in press; Rebuschat & Williams, 2006; Tagarelli, Borges Mota, & Rebuschat, 2011, 2014; Williams, 2005; Williams & Kuribara, 2008) often combine lexical information from one language (usually the L1 of the population being tested) and grammatical information from another. For example, Rebuschat and colleagues (Kim & Rebuschat, in prep; Rebuschat & Williams, 2006; Tagarelli et al., 2011, 2014) trained native speakers of English on a semi-artificial language consisting of English lexical items and German syntax. One advantage of semi-artificial languages is that the L1 lexical base obviates the need to learn novel word meanings, so participants can devote more cognitive resources to learning grammatical rules. This allows for faster acquisition, which is certainly a practical advantage. However, the ecological validity of this approach in testing L2 acquisition is questionable, as natural L2 acquisition involves learning of both lexicon and grammar.

Artificial languages (de Graaff, 1997; DeKeyser, 1995; Friederici et al., 2002; Hudson Kam & Newport, 2005, 2009; Morgan-Short, Finger, Grey, & Ullman, 2012; Morgan-Short,
Sanz, Steinhauer, & Ullman, 2010; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Newman-Norlund, Frey, Petitto, & Grafton, 2006; Opitz & Friederici, 2002; Yang & Givón, 1997), on the other hand, are typically composed of nonce words and phrase structure rules that are consistent with natural language rules. They can be spoken and comprehended, and because they are simplified models of natural languages, generally containing small lexicons (typically ranging from 10 to 100 words) and only a few target structures, they can be learned to high proficiency in a very short time. Behavioral evidence suggests that artificial languages can be used to train learners on lexicon and grammar within hours (Morgan-Short, Finger, et al., 2012; Morgan-Short et al., 2010; Morgan-Short, Steinhauer, et al., 2012; Yang & Givón, 1997), and that like natural languages, the success of learning may be related to instructional conditions (de Graaff, 1997; DeKeyser, 1995).

However, the extent to which artificial languages can inform us about L2 acquisition is still limited. These systems are artificial, and though they contain semantic and syntactic information, they rarely contain new phonological information. Mini-languages can address some of the shortcomings of artificial languages and other artificial linguistic systems.

1.2.2.1.5 Mini-language learning. A mini-language is a subset of a real natural language that maintains the structure, lexicon, and phonotactics of that language. Mini-languages have been used to “bridge artificial and natural language processing research” (Mueller, 2006, p. 247). Like artificial languages, a mini-language allows for the control of the amount and type of exposure, as well as control over the linguistic targets (e.g., word order, agreement). The structures in mini-languages are not merely consistent with natural language rules – they are taken directly from real languages. Artificial languages are an improvement over semi-artificial
languages in that they contain novel lexicon and grammar, but they still often follow the phonotactics of the target population (Friederici et al., 2002; Hudson Kam & Newport, 2005, 2009; Morgan-Short, Finger, et al., 2012; Morgan-Short et al., 2010; Morgan-Short, Steinhauer, et al., 2012). Mini-languages, on the other hand, include new phonotactic information and prosody, which are important components of L2 acquisition. In addition, mini-languages can start out simple and be made more complex ad infinitum by taking real examples from the full natural source language. Finally, native speakers of the source language should process the mini-language in the same way that they process their natural L1, which allows for the inclusion of an L1 control group in laboratory settings. This is especially important for studies using fMRI, which, unlike ERP, does not have known native language markers.

Studies on mini-languages demonstrate that these systems can be learned in a short amount of time (Batterink & Neville, 2013; Brooks & Kempe, 2012; Brooks, Kempe, & Donachie, 2011; Brooks et al., 2006; Kempe & Brooks, 2001, 2008; Kempe, Brooks, & Kharkhurin, 2010; Mueller, 2005, 2006; Robinson, 2005). They can be used to investigate the acquisition of specific features of a language, which include vocabulary and syntactic word order in most studies, as well as morphosyntactic features like noun declensions (Batterink & Neville, 2013; Brooks & Kempe, 2012; Brooks et al., 2006), case marking (Mueller, 2005, 2006; Mueller, Hahne, Fujii, & Friederici, 2005; Robinson, 2005), diminutive morphology (Brooks et al., 2011; Kempe & Brooks, 2001; Kempe et al., 2010), numerical and semantic classifiers (Mueller, 2005, 2006; Mueller et al., 2005), and locatives (Robinson, 2005).

Importantly, these studies demonstrate the extent to which mini-languages serve as a valid model for L2 acquisition. A series of studies on a miniature version of Russian by Brooks,
Kempe, and colleagues underscores the importance of studying multiple domains of language. By varying the number of lexical items presented over a fixed time frame prior to grammar learning, they found that when the mini-language included fewer lexical items and, as a result, more exemplars of each item, learners more successfully acquired morphosyntactic rules (Brooks et al., 2006). Conversely, learners who were better at discovering inflectional patterns on nouns were shown to be better at acquiring and retaining the meaning of lexical items (Brooks & Kempe, 2012; Brooks et al., 2011). Mini-language studies have also added to the literature on individual differences in L2 acquisition by showing a relationship between cognitive abilities, such as IQ and working memory, and learning outcomes (Brooks & Kempe, 2012; Brooks et al., 2006; Robinson, 2005).

Behavioral research on both natural language and artificial linguistic systems makes it clear that L2 acquisition is a complex process, and there are many factors that contribute to the difficulty of adult L2 acquisition. Certain features are harder to acquire than others, and while it is useful to look at language domains and learning processes in isolation, the interaction between these factors cannot be ignored. However, while the studies above provide behavioral evidence for the mechanisms involved in L2 processing, they only allow us to make inferences about the neural systems underlying L2 acquisition and use. Recent technology has allowed for the investigation of the neural bases of language, which brings us closer to answering questions about how the brain learns a language in adulthood. The next section focuses on the electrophysiological bases of L2 learning and processing.

1.2.3 Electrophysiological Evidence: Event-Related Potentials (ERPs). One widely used technique that directly measures the electrophysiological neural underpinnings of L2
processing is ERPs. ERPs are scalp-recorded electrical potentials of brain activity that are time-locked to the presentation of target stimuli, such as words, pictures or sounds. They have extremely sensitive temporal resolution, and can therefore provide online measures of language processing and comprehension in real-time. ERPs have been shown to distinguish different types of language processing in L1 (for reviews, see Kaan, 2007; Steinhauer & Connolly, 2008). Lexical/semantic processing difficulties in L1 have been shown to elicit N400s, which are central/posterior bilaterally distributed negativities that peak around 400 ms after the onset of a word (Bowden et al., 2013; Friederici, Cramon, & Kotz, 1999; Hahne, 2001; Hahne & Friederici, 2001; Kutas & Hillyard, 1980; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006). This component may be related to declarative memory (Ullman, 2001a). Syntactic processing difficulties, on the other hand, sometimes elicit a biphasic pattern of early (150 – 500 ms) left-to-bilateral anterior negativities (LANs; Bowden et al., 2013; Friederici, 1995; Friederici, Pfeifer, & Hahne, 1993); but see Hagoort & Brown, 1999; Neville, Nicol, Barss, Forster, & Garrett, 1991; Osterhout, Bersick, & McLaughlin, 1997; Osterhout & Holcomb, 1992, 1993) and P600s, which are late centro-parietal positivities peaking around 600 ms (Bowden et al., 2013; Hahne & Friederici, 2001; Kaan, Harris, Gibson, & Holcomb, 2000; Osterhout & Holcomb, 1992; Steinhauer & Connolly, 2008). LANs may be related to procedural memory (Hahne, 2001; Hahne & Friederici, 2001; Kaan et al., 2000; Osterhout & Holcomb, 1992; Ullman, 2001a, 2004). Syntactic violations have also been shown to elicit late anterior negativities, or sustained bilateral anterior negativities in the 600 – 2000 ms time window (Martín-Loeches, Muñoz, Casado, Melcón, & Fernández-Frías, 2005). These L1 ERP
components, particularly the N400 for semantic violations and the P600 for syntactic violations, provide a robust basis against which L2 processing can be compared.

In recent years, the electrophysiological basis of L2 has become more widely understood due to ERP research in this area. Importantly, ERPs have revealed effects of L2 acquisition in the absence of, or prior to, behavioral effects (J. McLaughlin et al., 2004; Shestakova, Huotilainen, Čeponiene, & Cheour, 2003; Tokowicz & MacWhinney, 2005). General trends in ERP research have shown that lexical/semantic processing difficulties in L2, like in L1, elicit N400s, even after minimal exposure to the L2 (Bowden, 2007; Hahne, 2001; Hahne & Friederici, 2001; J. McLaughlin et al., 2004; Moreno & Kutas, 2005; Steinhauser et al., 2009; Ullman, 2001a; Weber-Fox & Neville, 1996). In contrast, differences between L1 and L2 have been found in the processing of syntactic difficulties, especially at lower proficiency levels. At low proficiency L2, (morpho)syntactic violations sometimes elicit no early negativity at all (Hahne & Friederici, 2001; Ojima, Nakata, & Kakigi, 2005), or they elicit N400 or N400-like posterior negativities similar to those elicited for semantic violations in L1 and L2 (Bowden, 2007; Osterhout et al., 2008, 2006; Weber-Fox & Neville, 1996). At high proficiency L2, and at low proficiency in one study (Osterhout et al., 2006), (morpho)syntactic violations elicit P600s or P600-like late positivities (Bowden, 2007; Dowens, Vergara, Barber, & Carreiras, 2010; Isel, 2007; Osterhout et al., 2008; Rossi, Gugler, Friederici, & Hahne, 2006; Steinhauser et al., 2009; Weber-Fox & Neville, 1996), sometimes in conjunction with a LAN (Dowens et al., 2010; Hahne, Mueller, & Clahsen, 2006; Rossi et al., 2006; Steinhauser et al., 2009).

Most of the L2 ERP research investigates L2 processing in a single group of learners either at low or high proficiency, compared to a native speaker control group. Such studies
cannot capture learning-related changes in the brain, and therefore leave gaps in our understanding of the trajectory of L2 acquisition. In a cross-sectional design, Bowden et al. (2013) probed proficiency-based neural signatures by testing low proficiency L2 learners of Spanish, high proficiency learners, and native speakers of Spanish. This design revealed differences between low proficiency speakers and the other two groups in grammatical processing, but not in lexical/semantic processing, and striking similarities between the high proficiency and native speaker groups in grammatical processing. However, the use of different participants for low and high proficiency limits the extent to which the two groups are comparable.

Longitudinal designs are more desirable because they decrease between-subjects variability and allow for the observation of actual changes over time in the same group of subjects, therefore providing a true index of learning. Osterhout and colleagues (J. McLaughlin et al., 2004; Osterhout et al., 2006) have used longitudinal designs to test ERP effects in L2 learners. McLaughlin et al. (2004) followed native speakers of English through their first semester of French instruction, testing them on prime-target pairs of letter strings three times throughout the semester: first after about 14 hours of instruction, second after about 60 hours of instruction, and finally after about 140 hours of instruction. After only 14 hours, French pseudowords elicited greater N400s than French words, even though behavioral performance was only at chance. After 60 hours, smaller N400s were found when a word was primed by a related word than when it was primed by an unrelated word. After 140 hours, the amplitude of the N400s for both effects approximated those of native speaker controls, though behavioral judgments remained poor. The authors concluded that the learners were able to extract
information about word form after about only 14 hours of instruction, and information about word meaning after only 60 hours. In a longitudinal study testing morphosyntax, Osterhout et al. (2006) again tested native English speakers in their first semester of French, this time after one month, four months, and eight months of instruction. They were tested on semantic violations, verb/person agreement violations, which were phonologically realized, and number agreement violations, which were not phonologically realized. Learners who were categorized as “fast learners” based on behavioral performance showed N400 responses to semantic and verb/person agreement violations at one month, which changed to a P600-like positivity for the verb/person agreement at four months. There were no differences in ERP components for number agreement at any time throughout the study.

Studies like the ones described above not only demonstrate the distinct electrophysiological underpinnings of lexical/semantic and morphosyntactic processing; they also provide intriguing snapshots of how processing changes over the course of learning. On the other hand, these studies highlight the difficulty of capturing the course of language acquisition with natural language learning. Bowden et al.’s (2013) cross-sectional study captures changes in neural processing from low to advanced proficiency, but it misses aspects of the learning process in between. Additionally, because this study examines two distinct groups of subjects, proficiency effects cannot be completely deconfounded from between-subjects effects. Osterhout et al. (2006) and McLaughlin et al. (2004) examined changes over time in one group of subjects, which solves the problem of between-subjects confounds, but these studies only investigate distinct points in the learning process over one or two semesters. They therefore only capture the
early stages of learning, and furthermore only capture snapshots within those stages. Artificial linguistic systems provide an opportunity to investigate longitudinal learning to high proficiency.

1.2.3.1 Artificial linguistic systems. The dearth of longitudinal studies probing the electrophysiology of L2 acquisition has prompted studies using artificial linguistic paradigms. In general, results from these paradigms are similar to those from natural languages.

1.2.3.1.1 Word learning. Studies of word learning in controlled laboratory settings have reliably shown that recently learned words elicit N400s in incongruent contexts (Balass, Nelson, & Perfetti, 2010; Batterink & Neville, 2011; Frishkoff, Perfetti, & Collins-Thompson, 2010; Perfetti, Wlotko, & Hart, 2005), even after just one presentation of the word (Borovsky, Kutas, & Elman, 2010). Additionally, the P600 component has been shown to distinguish newly trained words from familiar words or untrained words in several word learning studies (Balass et al., 2010; Batterink & Neville, 2011; Frishkoff et al., 2010; Perfetti et al., 2005), which may index a familiarity effect and encoding for words presented in these experimental contexts. Evidence suggests that both the N400 and P600 effects may be modulated by the extent to which particular words are successfully learned (Balass et al., 2010; Batterink & Neville, 2011) or by individual differences in reading comprehension abilities (Balass et al., 2010; Perfetti et al., 2005).

1.2.3.1.2 Word segmentation. ERP evidence suggests that word segmentation also relies on the same mechanisms as those involved in natural language, which change over the course of learning. An N400 effect has been observed in word segmentation after as little as one minute of training, with the amplitude of this component increasing over time (Cunillera et al., 2009; De Diego Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Lévi, 2007; Sanders, Ameral, & Sayles, 2009). As exposure continues to increase, the N400 amplitude has been shown to decrease,
which has been suggested to reflect two separate processes: segmenting speech as the N400 increases, and recognizing already-segmented words as the component decreases (Cunillera et al., 2009). Other components have also been found when testing words presented in these paradigms, including N100, which may index word onset effects (Sanders et al., 2009), as well as more rule-based components like the P2 and P600 (De Diego Balaguer et al., 2007). This suggests that word segmentation may involve both semantic and grammatical processes, and that these processes change over the course of learning.

1.2.3.1.3 Artificial grammar learning. Artificial grammars have been used as models for language learning and processing for years, and ERP patterns suggest that this is warranted. Syntactic violations in artificial grammars have been shown to elicit similar biphasic early negativities and late positivities to those found in natural language violations, especially for finite-state grammars, after only a brief period of training (Bahlmann, Gunter, & Friederici, 2006; Friederici et al., 2002). However, in some cases, only a P600 is observed (Hoen & Dominey, 2000; Lelekov-Boissard & Dominey, 2002; Tabullo et al., 2011), suggesting that some artificial grammars might be better models of natural language than others. In any case, it is not clear that any artificial grammar learning studies have investigated changes over time in the electrophysiological signatures of L2, even though this could easily be done.

1.2.3.1.4 Artificial language learning. Artificial language paradigms allow for the investigation of changes in processing of semantic and syntactic incongruences in the same group of subjects over a learning period. Findings from these studies provide some of the best insights on the course of language learning, showing artificial language models elicit similar ERP components to those found in natural language processing. Friederici et al. (2002) trained
adults on a simple artificial language, BROCANTO, until they reached high proficiency in the language (95% accuracy). Syntactic violations elicited the biphasic LAN/P600 response found in L1 and high proficiency L2 speakers. Morgan-Short and colleagues (Morgan-Short, Finger, et al., 2012; Morgan-Short et al., 2010; Morgan-Short, Steinhauer, et al., 2012) used a modified version of BROCANTO (BROCANTO2) to investigate the effects of training (i.e., explicit, in which the rules of the language were explained to the learners, and implicit, in which meaningful phrases and sentences were presented to the learners with no instruction) on L2 neurocognition at low and high proficiency. Behaviorally, learners in the explicit and implicit training conditions performed similarly on a grammaticality judgment task (GJT). However, the ERPs elicited by syntactic violations were different across training conditions. In the implicit training condition only, syntactic violations elicited N400s at low proficiency and a biphasic LAN/P600 pattern at high proficiency, similar to what has been found in natural language learning. Morphosyntactic violations elicited different ERP components for the implicit and explicit training conditions, but neither approximated the natural-language-like LAN/P600 response. When learners were tested again three to six months after the initial training period (Morgan-Short, Finger, et al., 2012), behavioral performance on syntactic processing did not change for either group, but ERP effects did. Specifically, for both groups at retention, the anterior negativity appeared earlier, was more robust, and was more left-lateralized, and the P600 was more robust. In other words, both groups showed more native-like ERP patterns for syntactic processing after several months had passed without intervening training. Together, these studies show that adult learners can achieve native-like processing in conjunction with high performance in an L2, but that the neural signatures of language processing change over the course of learning. Furthermore, the BROCANTO2 studies
demonstrate that even after high proficiency is achieved, a shift in processing mechanisms can occur.

1.2.3.1.5 Mini-language learning. Research on the electrophysiology of mini-languages adds to the work done on artificial languages, but with a more natural language system. Mueller and colleagues (Mueller, 2005; Mueller et al., 2005) used a miniature version of Japanese, called Mini-Nihongo, to assess word category, case, and classifier processing in native and non-native (L1 German) speakers of Japanese. In the first study (Mueller et al., 2005), non-native speakers were trained on the language for an average of 7.2 hours, until they reached high proficiency. The training was followed by a GJT, during which ERPs were recorded. Behavioral results indicated a learning effect on all aspects of language being tested: native and non-native speakers did not differ in performance on correct sentences, word category violations, or classifier violations, but non-native speakers performed significantly worse than non-native speakers on case violations, though their accuracy was still high (84%). For Japanese native speakers, word category violations elicited a broad anterior negativity between 100 and 300 ms, followed by a P600. Case violations elicited an early negativity between 100 and 300 ms, an N400, and a P600, and classifier violations elicited a late LAN (500 – 800 ms). For non-native speakers, training resulted in a more centrally-distributed N400 followed by a P600 for word category violations and a P600 only for case violations. There was no observed ERP effect of classifier violations in non-native speakers. The authors conclude that high proficiency non-native speakers show striking similarities to native speakers in the ERP components elicited in response to violations, though the non-native speakers seem to rely less on automatic processes (as evidenced by the lack of early negativity in response to case violations) and more on prosodic cues, as indicated
by the more centrally-distributed N400 in response to word category violations. Importantly, Mueller (2006) notes that the differences between this study and Friederici et al.'s (2002) study on the ERP components to violations in BROCANTO suggest that the findings in the artificial language study are not completely generalizable to natural language input.

In a follow-up study focusing on word order and case (Mueller, 2005), the same participants in the original study underwent additional training in Mini-Nihongo, which resulted in improved performance on case violations (i.e., non-native speakers no longer performed worse than native speakers). ERP components for native-speakers were, unsurprisingly, similar to those found by Mueller et al. (2005). ERP components did not change for non-native speakers on word category violations after additional training. However, on case violations, a frontally distributed N400 was elicited for sentences with canonical word order. This demonstrates that additional training did not result in more native-like processing on word order, but it did result in more native-like processing for case. Taken together, these studies suggest a role for prosodic processes and controlled syntactic processes in L2 learners, the former of which could only be investigated because this was a mini-language, which contained natural connected speech.

In a longitudinal study, Davidson and Indefrey (2009) used a mini-language learning design to teach native speakers of Dutch about gender and adjective declension in German. They recorded ERPs during pretest, training, and posttest phases of the experiment. The pretest was a GJT in German. This was followed by a learning task, in which participants studied vocabulary items, completed tests regarding gender and Dutch-German noun translation, and read a description of the target grammatical rules. In the training task, participants performed a GJT and received feedback (correct/incorrect) on their responses. The posttest, which occurred one week
after the other phases, consisted of a GJT without feedback, like the pretest. Behavioral results show that learners improved over the course of the experiment. In the training and posttest phases, but not the pretest, both native and non-native speakers showed P600-like effects for declension violations. Gender violations did not elicit ERP components in learners at any phase of the experiment, whereas they elicited P600 effects in native speakers. This study reveals neural and behavioral changes over the course of learning in a short time period, and differences between L1 and L2 processing, but it only looked at snapshots in this time course, rather than the full trajectory of acquisition.

While Davidson and Indefrey’s (2009) study provides intriguing results stemming from a longitudinal design, there are some issues that make it a marginal example of a mini-language study. Although this study used a mini-language based on four German adjectives and 40 nouns to teach non-native speakers about a new morphosyntactic rule, the learners in this study had all had previous coursework in German in high school, so it straddles the fence between a natural L2 study and a mini-language study. Furthermore, the authors note that the similarities between Dutch and German should have allowed the learners to adapt quickly to the experiment, and that they intentionally chose to investigate similar languages to allow for fast learning. While such a choice was likely effective in promoting faster learning, it limits the generalizability of this study to the acquisition of more typologically distant languages.

ERP evidence from natural languages and artificial linguistic systems suggest distinct patterns for lexical/semantic and grammatical processing, with lexical/semantic processing evoking N400s and grammatical processing evoking biphasic early negativities and late positivities, or P600s. Additionally, there appear to be shifts in the neural mechanisms involved
in learning, particularly for grammar. However, while ERPs are extremely useful for understanding the temporal aspects of language processing, they give extremely vague indications, if any, of where language is being processed in the brain. For this, we must turn to hemodynamic neuroimaging research, which uses techniques such as fMRI and positron emission topography (PET) to measure activity in the brain based on blood-oxygenation levels.

1.2.4 Functional Neuroanatomy: PET and fMRI. Studies using fMRI and PET have revealed different patterns of activation for lexical/semantic and grammatical processing in L1 (for a review see Hasson & Small, 2008). Lexical/semantic processing has been linked to activation in temporal and temporo-parietal regions, including the MTL (Friederici, Kotz, Werheid, Hein, & Yves, 2003; Illes et al., 1999; Kuperberg et al., 2000; Newman, Pancheva, Ozawa, Neville, & Ullman, 2001b). The selection, retrieval, and integration of lexical/semantic knowledge seem to be subserved by BA45 and BA47, which include the pars triangularis of the IFG and surrounding areas (Dapretto & Bookheimer, 1999; Illes et al., 1999; Poldrack et al., 1999). L1 (morpho)syntactic processing appears to rely on BA44, or the frontal operculum of the IFG, as well as the superior temporal gyrus (STG) and the basal ganglia (Dapretto & Bookheimer, 1999; Friederici et al., 2003; Newman, Pancheva, Ozawa, Neville, & Ullman, 2001a).

Hemodynamic neuroimaging studies of L2 have begun to shed light on the neural structures involved in L2 processing (for comprehensive reviews see (Abutalebi, 2008; Grey, Tagarelli, Turkeltaub, & Ullman, 2013; Indefrey, 2006; Kotz, 2009; Morgan-Short & Ullman, 2012; Stowe & Sabourin, 2005; Tagarelli, Grey, Ullman, & Turkeltaub, 2012; Ullman, 2005; Ullman, Tagarelli, Grey, & Turkeltaub, in prep). Similarly to the computational and
electrophysiological evidence discussed above, the neural structures and activation patterns underlying lexical/semantic processing tend to be the same in L1 and L2 (Chee, Tan, & Thiel, 1999; Klein, Milner, Zatorre, Zhao, & Nikelski, 1999; Xue, Dong, Jin, Zhang, & Wang, 2004), though some studies have found greater activation in L1 compared to L2 (Pillai et al., 2004), or vice versa (Perani et al., 2003). The results for grammatical processing are less clear. Some studies have found that grammatical processing in L2 involves more temporal lobe involvement than in L1, bilaterally, and furthermore, that temporal lobe involvement is greater for later learners or those with less exposure than for earlier learners or those with more exposure (Dehaene et al., 1997; Perani et al., 1996, 1998). The extent to which the frontal lobe is involved in L2 processing is also somewhat unclear. Some studies have found that frontal activation is similar in L1 and L2 processing (Dehaene et al., 1997; Perani et al., 1996, 1998), while others have found greater frontal activation in L2, but in different regions of the frontal lobe, including BA44, 45, and 47 (Golestani et al., 2006; Hasegawa, Carpenter, & Just, 2002; Nakai et al., 1999; Rüschemeyer, Fiebach, Kempe, & Friederici, 2005; Wartenburger et al., 2003).

Because the L2 PET and fMRI literature is somewhat inconsistent, we conducted an Activation Likelihood Estimation (ALE; Eickhoff et al., 2009; Eickhoff, Bzdok, Laird, Kurth, & Fox, 2012; Turkeltaub et al., 2012; Turkeltaub, Eden, Jones, & Zeffiro, 2002) meta-analysis of lexical/semantic and grammatical PET/fMRI studies of L1 and L2 (Grey et al., 2013; Tagarelli et al., 2012; Ullman et al., in prep). These analyses suggest that lexical/semantics yields similar activation patterns in L2 and L1 (e.g., BA45/47, temporal pole, STG), though with more extensive and bilateral activation in L2. In contrast, there seems to be overlap as well as differences between L2 and L1 grammar. Activation in neocortical regions, including BA44 and
BA6, was found for both L2 and L1 grammar, again more extensively and bilaterally in L2. Additionally, L2 grammar elicited basal ganglia activation, in particular in the caudate nucleus. Activation in this area was not found in L1 grammar or in L1 or L2 lexical/semantics. Intriguingly, L1 and L2 lexical/semantics and L2 grammar showed activation in a posterior parietal region that has also been found to be larger in L2 learners than in monolinguals (Mechelli et al., 2004), correlates with vocabulary size in monolinguals (Lee et al., 2007), and has been implicated in declarative memory (Wagner, Shannon, Kahn, & Buckner, 2005). This area was not activated in L1 grammar. This meta-analysis provides quantitative evidence to suggest that L2 grammar may rely on neural substrates that also underlie lexical/semantics, vocabulary, and declarative memory, in addition to those that underlie L1 grammar.

1.2.4.1 Artificial linguistic systems. Research probing the functional neuroanatomy of language learning using artificial systems has made important contributions to this field by uncovering the structures underlying processing in language domains in isolation. Perhaps more importantly, several of these studies have uncovered the neural correlates of language learning in longitudinal designs.

1.2.4.1.1 Word learning. Laboratory studies have identified several brain areas that seem to be involved in the acquisition of new lexical items (for a review, see Laine & Salmelin, 2010). These include the hippocampus, parahippocampal gyrus, and other MTL regions in early stages of learning (Breitenstein et al., 2005; Mestres-Missé et al., 2008, 2010; Raboyeau et al., 2004), as well as the LIFG (Mestres-Missé et al., 2008, 2010; Raboyeau et al., 2004, 2010), the inferior parietal lobe (Breitenstein et al., 2005; Cornelissen et al., 2004), and the middle temporal cortex (Mestres-Missé et al., 2008, 2010; Raboyeau et al., 2004). Activation in the hippocampus,
fusiform gyrus, inferior parietal lobe, and temporal regions have been shown to correlate with behavioral performance, but the direction of the correlations have not been consistent (Breitenstein et al., 2005; Raboyeau et al., 2004, 2010).

The research on word learning in laboratory settings suggests that this is a relatively easy task for learners, and behavioral evidence indicates a reliance on phonological cues in the early phases of acquisition. Neural evidence corroborates this reliance, demonstrating activation in parietal areas associated with phonological encoding and processing. Areas related to learning, memory, and language have also been shown to be involved in novel word learning.

**1.2.4.1.2 Word segmentation.** FMRI studies of word segmentation are particularly useful because they allow for continuous scanning during learning, or exposure to the speech stream, and therefore can probe learning-related changes in the brain. This is accomplished by examining the relationship between accuracy or exposure and neural activation. In particular, these studies have shown a positive relationship between learning (i.e., accuracy and/or amount of exposure) and activation in bilateral caudate (Karuza et al., 2013; Scott-Van Zeeland et al., 2010), bilateral putamen (Karuza et al., 2013; Scott-Van Zeeland et al., 2010; but see McNealy, 2006), STG (McNealy, 2006; McNealy et al., 2011), and supramarginal gyrus (McNealy, 2006; Scott-Van Zeeland et al., 2010). The inferior parietal lobe (McNealy, 2006; Scott-Van Zeeland et al., 2010) and the LIFG (Karuza et al., 2013; Scott-Van Zeeland et al., 2010) seem to play important roles in this process, but the relationship to performance or exposure is less clear.

**1.2.4.1.3 Artificial grammar learning.** A comprehensive review of the functional neuroanatomy of artificial grammar learning is provided by Folia, Uddén de Vries, Forkstam, & Petersson (2010). The evidence suggests that many overlapping neural regions are in fact
activated during artificial and natural language syntactic processing. For example, Petersson et al. (2012) trained participants on a finite-state artificial grammar and acquired fMRI data during a grammatical classification task. There was activation in inferior and middle frontal regions bilaterally, which was centered on BA44 and 45, inferior parietal cortex, including BA39 and 40, middle and inferior occipital regions, including BA18 and 19, posterior mid-inferior temporal regions, including BA20 and 21, and the cerebellum and basal ganglia. These regions of activation are similar to those found in other studies investigating artificial grammar learning with fMRI (Forkstam, Hagoort, Fernandez, Ingvar, & Petersson, 2006; Petersson et al., 2004). However, despite some overlap, studies have shown important neural and functional distinctions between nonlinguistic and linguistic grammar acquisition (Fitch & Hauser, 2004; Friederici, 2004; Friederici et al., 2002; Opitz & Friederici, 2003).

One strand of research on linguistic artificial grammars trains learners on two types of artificial grammars with pseudowords: one has rules that follow natural-language universals, and the other has rules that do not (Musso et al., 2003; Tettamanti et al., 2002). In these studies, participants learned the rules to a high proficiency criterion. Musso et al. (2003) recorded fMRI during a GJT following training. They found increasing activation in BA45 as proficiency increased for rules following natural-language patterns, but not for the other rules. Tettamanti et al. (2002) found activation in BA44, again only for natural-language rules. They also found a relationship between activation and behavioral performance; there was greater activation in learners with the highest proficiency, as compared to learners with lower proficiency.

Opitz and Friederici’s (2003) investigated artificial grammar learning continuously using fMRI. In this study, participants viewed grammatical sentences in BROCANTO and were asked
to attempt to extract the rules of the artificial language. In the testing phase, participants performed a GJT on sentences that either did or did not conform to the rules of BROCANTO. Behavioral performance improved over time. The neuroimaging analysis revealed the recruitment of the hippocampus and the temporal cortex in early stages of learning, with activation in this area decreasing over the course of learning. Activation in BA44, on the other hand, increased over the course of learning, suggesting a shift during learning from the declarative memory processes typically involved in lexical/semantic processing to the native-like syntactic processing associated with procedural memory structures (Ullman, 2001c).

1.2.4.1.4 Artificial language learning. Only one study that the author is aware of investigated artificial language learning with fMRI (Newman-Norlund et al., 2006). Newman-Norland et al. (2006) performed fMRI scanning at low, medium, and high proficiency on learners of Wernicke's, a finite-state grammar containing meaningful words. They found a positive relationship between proficiency and activation in the LIFG (BA44), lingual gyrus, anterior cingulate cortex, premotor cortex, and putamen, among other regions. The STG (planum temporale) was also associated with learning, though patterns of activation in this area were modality-dependent. The fMRI findings from artificial grammars and artificial languages are consistent with those for L2 computations and electrophysiology, which also demonstrate a shift for syntactic processing over the course of L2 acquisition.

1.2.4.1.5 Mini-language learning. At this point, there are no studies, to the author’s knowledge, that have investigated learning of a mini-language, as defined here, using fMRI or PET. Note that while some studies refer to their linguistic systems as mini-languages, these are
usually artificial languages, and sometimes artificial grammars, by the definitions in this dissertation.

1.2.5 Summary of Evidence. In sum, studies on the acquisition of natural languages in adulthood reveal a distinction between the acquisition and processing of lexical/semantic and grammatical aspects of the L2, and further distinctions within the realm of (morpho)syntax. Learners acquire lexical/semantic features of the L2 faster and with greater ease than (morpho)syntactic features, which is unsurprising, considering the evidence that L2 lexical/semantic processing depends on what appears to be the same computational and neural underpinnings as L1 lexical/semantic processing. In contrast, L2 (morpho)syntactic processing relies on computations and structures that are different from L1 (morpho)syntactic processing, particularly at lower proficiencies. In fact, L2 (morpho)syntactic processing seems to depend on the mechanisms involved in L1 and L2 lexical/semantic processing in the early stages of acquisition, and shifts to reliance on those structures and computational mechanisms involved in L1 (morpho)syntactic processing as proficiency increases.

Research to date on the acquisition of real, full natural languages provides extremely important insights about L2 processing at different proficiency levels and a sturdy foundation from which further research can be launched. However, important questions remain unanswered. The few studies that have used longitudinal designs do provide evidence of a qualitative shift in the neurocognition of (morpho)syntactic processing, but they only provide snapshots of L2 processing over the course of learning. Unsurprisingly, no study has actually investigated the entire course of learning a natural L2 either behaviorally or neurally; the amount of time and effort that goes into learning a natural L2 makes such an experiment logistically impossible.

38
Furthermore, the amount, timing, and type of language exposure is difficult to control in studies that investigate L2 acquired in naturalistic contexts, thus creating potential confounds in the observed results. To address some of these confounds and limitations, researchers have turned to studies of artificial linguistic systems, which have examined L2 acquisition and processing in laboratory settings.

Overall, the current research suggests that the behavioral, computational, and neural processes involved in learning artificial linguistic systems are similar to those involved in native language processing and natural L2 processing, which indicates that such systems provide appropriate models for L2 acquisition (Mueller, 2006). Furthermore, artificial linguistic systems allow for the continuous longitudinal investigation of language learning, though there are few studies that take advantage of this. The handful of studies that have investigated changes in learning over time are informative, but the extent to which they can inform us about L2 acquisition is limited. These systems are still artificial, and though some contain semantic and syntactic information, they rarely contain new phonological information or even prosody, and thus still present language domains in isolation. At this point, there are few studies that investigate longitudinal and continuous learning at all, and none that investigate the behavioral and neural underpinnings of the course of L2 learning using a more natural linguistic system.

1.3 The Current Study

In this dissertation project, behavioral and neural (fMRI) data were collected over the course of acquisition of a mini-language, based on Basque, from zero knowledge to advanced proficiency. Through this design, it was possible to examine neural and behavioral changes in real-time during both training and testing with natural language stimuli. These changes were
observed both independently, as well as in relation to each other. Additionally, the incorporation of lexical/semantic, syntactic, and morphosyntactic targets in the design allowed for the investigation of differences in performance for different aspects of language.

1.3.1 Research questions. The research questions for this study are as follows:

1) How does behavioral performance change over the course of adult L2 acquisition?
2) Are the neural correlates of adult L2 acquisition the same as the neural correlates of L1 processing?
3) Does neural activity change over the course of adult L2 acquisition? If so, how?
4) Do changes in neural activity and behavioral performance differ according to aspects of language, specifically lexical/semantics, syntax, and morphosyntax?
5) Is there a relationship between behavioral and neural aspects of L2 learning? If so, does this change over the course of learning?

1.3.2 Hypotheses. The predictions from these research questions are derived from previous research and the three prominent neurocognitive theories of L2 neurocognition.

H1: Behavioral performance is expected to improve over time, for all aspects of language, as evidenced by increases in accuracy and decreases in reaction time (RT). Words and lexical/semantic information are expected to be learned quickly, followed by word order and morphosyntax, which is expected to be most difficult to learn. An asymptotic pattern of learning is predicted for all language domains, in which greater changes in performance will be seen in early stages of learning, with smaller and smaller changes as training progresses. Comprehension abilities are expected to be better than production.
H2: If L1 and L2 are fundamentally distinct, we would expect to find activation in areas that are not typically activated during L1 processing. We cannot directly test this in our study, as we do not have a measure of L1 processing for these subjects. However, results from our meta-analysis have revealed general patterns in L1 lexical/semantics and grammar with which we can compare observed mini-language activation in the proposed project (Grey et al., 2013; Tagarelli et al., 2012; Ullman et al., in prep). According to our meta-analysis, L1 lexical/semantics depends on areas of the LIFG, including BA44, 45, and 47, precentral gyrus, temporal pole, STG, and posterior parietal cortex. L1 grammar depends on left BA44/45, STG, and the cerebellum. Therefore, a lack of any substantial overlap with these areas during all phases of mini-language learning could lend some tentative support to this hypothesis.

Both the Convergence Hypothesis and the DP model predict involvement of the LIFG and the basal ganglia in L2 learning. Within the IFG, the DP model especially predicts activation in BA6, 44, 45, and 47. The DP model also predicts involvement of the hippocampus, medial temporal lobes, and temporal cortex, and the Convergence Hypothesis predicts involvement of temporo-parietal cortices. The involvement of these areas should be modulated by language domain and proficiency, as described in H3 and H4.

H3 and H4: The Fundamental Difference Hypothesis makes no specific claims about how neural and behavioral correlates might change as a function of exposure and proficiency. However, this hypothesis is based on the assumption that age is far more important than proficiency in determining the neural correlates of L2. Age is relatively constant in this study, so no changes would be expected according to the Fundamental Difference Hypothesis.
The convergence hypothesis makes more specific predictions according to level of proficiency and language domain. In early stages of learning, lexical/semantics should show broad activation in temporo-parietal regions and the LIFG, areas that are also activated in L1, but to a lesser extent. Activation should also extend to the PFC and basal ganglia, which are thought to be involved in the recruitment of controlled processes necessary for using a “weaker” L2 system. Over the course of lexical/semantic learning, PFC and basal ganglia activation should decrease, and activation in temporo-parietal regions and LIFG should be less broad. For grammar, early learning should depend on broad activation in the LIFG and extended activation to the PFC and basal ganglia. Because AoA is thought to be more important than proficiency in determining the neural representation of L2 grammar, and all subjects in the proposed study are late L2 learners, this theory does not necessarily predict any changes in grammar-related activation over the course of learning. However, this view does not entirely discount the role of proficiency in L2 grammar representation, so decreases in PFC and basal ganglia activation overall, and a decrease in the extent of LIFG activation, would still be somewhat consistent with this hypothesis.

The DP Model predicts an initial reliance on MTL structures, especially the hippocampus and surrounding areas, for both lexical/semantic and (morpho)syntactic learning. As learning progresses, lexical/semantics should rely less on the MTL and more on neocortical structures, including the temporal cortex and LIFG, specifically BA45 and 47, and potentially the posterior parietal cortex. Activation for grammar should also decrease in the MTL, and should increase in neocortical regions that subserve declarative memory (BA45, 47, temporal cortex, parietal cortex) and procedural memory (BA44 and 6). Activation is also predicted in the basal ganglia,
especially the caudate nucleus. Because agreement is more difficult to learn than word order, these changes in the neural correlates of grammar may occur earlier and more completely for the latter language domain. At very high proficiency, when grammar skills are automatized, declarative memory structures and the basal ganglia may no longer be activated.

H5: Research questions 3 and 4 conflate the issue of exposure and proficiency based on the assumption that as exposure increases, so does proficiency, and thus changes over the course of learning are related to changes in both of these variables. Research question 5 attempts to tease these variables apart by probing not just changes in activation over time, but differences in activation as a function of proficiency. In other words, is activation in a particular run of exposure related to accuracy in that run? Is there a difference in neural activation between subjects who are better or poorer learners? None of the neurocognitive models discussed here makes separate predictions for exposure and proficiency – the same neural changes are expected as learners gain more exposure to the language as when learners become more proficient in the language. However, exposure and proficiency are not linearly or uniformly related, as evidenced by the high degree of variability in success among L2 learners, even when they have the same amount and type of exposure. Therefore, answering this research question may reveal informative differences between exposure and proficiency in neural development during adult language learning.

The current study aims to answer the research questions and test the hypotheses outlined above. The proposed methodological approach is outlined in detail in the following section.
Chapter 2: Research Methods and Design

2.1 Participants

Seventeen monolingual native-speakers of English participated in this study. Of the 17, one was excluded for failing to follow directions, and one was unable to complete the study. The 15 remaining participants (7 female) ranged in age from 18 to 26-years-old ($M = 20.9$, $SD = 2.79$) and were right-handed (scores ranging from 78.9 to 100% on the Edinburg Handedness Inventory; Oldfield, 1971). They had normal hearing, normal or corrected to normal vision, and no known history of neurological, learning, or psychiatric disorders, neural injury or concussions, or drug or alcohol dependence. They were all enrolled in college or graduate school at the time of participation, or had completed at least four years of study at a post-secondary institution.

Of the participants who completed the study, behavioral results for all 15 were analyzed. Due to an MRI scanner malfunction, one participant’s data was completely excluded from fMRI analyses, and one participant’s data was excluded for any analyses involving Day 2. The decision was made to analyze all viable data, and therefore analyze data from different numbers of participants for different analyses, in order to maximize power for each analysis. This solution was more feasible than collecting more data, as data collection for each subject involved approximately 15 hours of participation and substantial costs for the use of the MRI scanner. Because of the challenges involved in collecting fMRI data and the development of appropriate statistical approaches for dealing with small subject groups, studies with sample sizes of 15 or fewer participants are very common in fMRI studies of L2 (Abutalebi et al., 2013; Chee, Hon, Lee, & Soon, 2001; De Bleser et al., 2003; Ding et al., 2003; Grey et al., 2013; Halsband,
Krause, Sipilä, Teräs, & Laihinen, 2002; Hernandez, Hofmann, & Kotz, 2007; Hernandez & Meschyan, 2006; Luke, Liu, Wai, Wan, & Tan, 2002; Luke et al., 2002; Meschyan & Hernandez, 2006; Ruschemeyer et al., 2005; Sakai et al., 2004; Saur et al., 2009; Stein et al., 2009; Tagarelli et al., 2012; Tatsuno & Sakai, 2005; Vingerhoets et al., 2003; Wartenburger et al., 2003; Yokoyama et al., 2006, 2009).

2.1.1 Language background. All participants had some exposure to second languages, but none reported being fluent, or having ever been fluent, in a language other than English. Fluency is conservatively defined here as the ability to converse freely in a natural, unstructured linguistic environment with minimal communication barriers, but not necessarily in a native-like way (Brumfit, 1984; Chambers, 1997; Crystal, 1987; Kormos & Dénes, 2004). If participants had any exposure to languages containing grammatical properties integral to Mini-Basque (e.g., Japanese, Latin, Turkish, Finnish, German), they met the following criteria for those languages: no more than 4 years of classes before college or 2 years in college, no more than 2 weeks of immersion, and no exposure over the past year. No participants had any training in Basque.

2.2 Materials

2.2.1 The mini-language: Mini-Basque. The target language in this study was a mini-language, operationalized as a subset of a real natural language that is simple enough to be learnable over the course of several hours in a laboratory setting. The mini-language was based on Basque, and the target language will henceforth be referred to as Mini-Basque. Using a subset of a natural language allows for the control for many factors, which are outlined in Section 2.2.1.1.2.
2.2.1.1 The source language: Basque. The mini-language used in this study is a subset of Basque. In this section, I outline the process used to determine the source language for this experiment, provide a brief overview of the Basque language, discuss the steps involved in designing Mini-Basque, and detail the characteristics of Mini-Basque.

2.2.1.1.2 Choosing a source language. Several selection criteria were used to determine the source language for the mini-language, some of which pertain specifically to the features of the language and some of which address practical concerns. The criteria and the related reasons are listed below. According to these criteria, the language must:

1) be simple enough for native speakers of English to learn to high proficiency over a short period of time, to allow for continuous training in the MRI scanner.

2) contain at least some relatively simple and quickly learnable morphological features, while still maintaining enough complexity to allow for the investigation of the acquisition of morphosyntax, such as number or gender agreement.

3) have a phonotactic system that does not differ vastly from English phonology. Because phonological acquisition is not of primary interest in this study, phonology should not be a factor that significantly impedes learning of words, morphosyntax, and syntax. This criterion excluded tonal languages, such as Mandarin, and any languages with sounds that are difficult for English speakers, like the retroflex consonants that exist in many Indian languages, such as Hindi and Gujarati. It should be noted, however, that languages that have complex morphology or difficult phonology for some structures were not completely discounted, as long as simple
enough structures were available in the language and particularly difficult phonemes could be avoided.

4) have a different word order from English (i.e., not subject-verb-object; SVO). This is because word order acquisition is of particular interest in this study, and using a language with a different word order from English helps to avoid the confound of positive L1 transfer effects on L2 word order processing. The source language could include SVO constructions, but other word orders should be commonly used in the language.

In addition to the linguistic criteria listed above, several logistical concerns were taken into consideration when determining the source language. In order to maximize the size of the potential participant pool, the source language could not be one that many learners in the U.S. may be exposed to on a regular basis. For example, Spanish, which is offered as a foreign language by 88% of elementary schools and 93% of secondary schools with language programs in the U.S. (Rhodes & Pufahl, 2009) and spoken at home by 12% of the U.S. population (Shin & Kominski, 2010), was not a candidate for a source language in this project. In contrast, because L1 speakers were needed to act as native speaker informants, record stimuli, and eventually participate as control participants in future studies, the source language could not be obscure or “dead” (e.g., Latin).

In searching for a source language that most closely met the above criteria, all languages listed in the World Atlas of Language Structures (WALS; Dryer & Haspelmath, 2013) were considered. At the time of the search, WALS contained 2,678 world languages organized by features. Though this is not an exhaustive list of the world’s languages, any languages not listed
in WALS are likely to be obscure and not well documented, and therefore would not meet our logistical criteria. The Ethnologue (Lewis, 2009) was used to supplement data from WALS. Languages were eliminated based on many linguistic features, including SVO word order, uncommon or difficult phonetic/phonological features (e.g., clicks, pharyngeal consonants, tone, large phonemic inventory), and some morphological characteristics, and on several practical considerations, including number of speakers, availability of L1 speakers, and available resources. Information about practical considerations was largely obtained through Google searches and correspondence with language experts.

2.2.1.3 Features of Basque. Following the above criteria, the source language chosen for this project was Basque. Basque is the last remaining pre-Indo-European language in Western Europe (Trask, 1997), and is spoken as a native language by slightly more than 700,000 people, mainly in areas of northern Spain and southwestern France (Dryer & Haspelmath, 2013; see Figure 1).

Basque morphology as a whole is actually quite complex, but it was nonetheless possible to isolate a sentence structure that is relatively simple, but still includes morphosyntactic agreement across short and long distance dependencies (see Section 2.2.1.3). This involves phonologically realized number marking, which is important for two reasons. First, singular and plural items can be distinguished in the aural mode, which is a more natural way to present language data. Second, the use of number as a feature in the mini-language allows for the creation of more sentences from fewer lexical items, which was important for creating a sufficient number of stimulus sentences in the mini-language. Basque has a small phonemic inventory, and its phonotactics do not differ vastly from English, so phonology was not expected
to substantially impede learning. Though Basque is a free word order language, its canonical word order is Subject-Object-Verb (SOV; Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells, 2009), which is different from English word order (SVO).

Figure 1. The Basque Country of Europe. © Zorion, CC-BY-SA, Wikimedia Commons.

Practically, Basque is a language isolate (de Rijk, 2008), and it is rarely taught or spoken in the U.S.² Its words and grammar are distinct from English, as well as from many other commonly taught languages in the U.S., such as Spanish and French. None of the participants screened for this study had any exposure to Basque, and very few were excluded based on exposure to languages with similar structures. Although Basque is not a commonly spoken language, it is spoken widely enough and has a strong academic following, which made it possible to establish contact with native speakers for consultation and stimulus recording.

² There is actually a distinct but rather small Basque-American population centered mainly in the western U.S. (Douglass, 2013), but this was not anticipated to hinder the subject acceptance rate for this study, and indeed, did not.
2.2.1.2 Designing Mini-Basque. Mini-Basque was developed with the help of several Basque-speaking collaborators and informants, and refined through pilot testing. The words and structures in Mini-Basque were created based on several resources on the Basque language (Aulestia & White, 1992; de Rijk, 2008; Laka, 1996), as well as suggestions for development and approval for the final version from two native Basque speaking informants at Georgetown University and a native Basque speaking professor of linguistics at the University of the Basque Country in Spain.

Regarding sentence structure, it was determined that sentences should contain transitive verbs with an agent and patient. To allow for reversibility of sentences, only animate nouns were considered. Specifically, common animals were chosen because their prototypes are likely to be more universally recognized across cultures than types of people (i.e., a cow in Spain looks more or less like a cow in the U.S., whereas a mailman might dress differently in different countries). Additionally, by not using people, the issue of biological gender could be more easily avoided (Basque does not have grammatical gender). Adjectives were also included to allow for variation in the complexity of the sentences. Color adjectives were chosen because these were most clearly representable in images. All words had to be concrete and representable in images, and an attempt was made to use words with the fewest possible syllables.

A verb norming study was conducted in order to choose four Basque verbs that best met the following criteria: 1) transitive, 2) takes animate agent and patient as arguments, 3) concrete, 4) can be represented in pictures, and 5) fits in the present perfective structure (see Section 2.2.1.3). From a list of the 400 most frequent transitive verbs in English (Davies, 2008), any that could not take an animate agent and patient or were not concrete were excluded. The remaining
forty verbs were translated into Basque using dictionaries, and checked with a native speaker. Some verbs have multiple translations in Basque, so this yielded a list of 46 concrete transitive verbs that take animate arguments. A Google Form survey was created, in which participants were asked to rate how good each sentence sounded in Basque (1 = very bad; 6 = very good), when each verb was inserted in the following sentence: Behiak txakurra ___________ du. (trans.: The cow (has) ___________ the dog.) Twelve native Basque speakers living in the Basque Country of Spain responded to the survey. Fourteen verbs were rated at least 5 or higher. Of those 14, the four verbs that were chosen for Mini-Basque, *ikusi* (6.00 ± 0; trans.: see), *biltzatu* (5.50 ± 1.0; trans.: push), *usaindu* (5.33 ± 1.37; trans.: smell), and *miazkatu* (5.25 ± 1.48; trans.: lick), were internally judged to be the most clearly representable in pictures, and we evenly distributed according to whether the patient of the verb was completely passive (as in see and smell), or somewhat reactive (as in lick and push).

Images were piloted to test the nameability of items and to verify that the images represented the intended concepts for both Basque and English native speakers. Images were drawn based on Snodgrass and Vanderwart’s (1980) standardized set of pictures, and modified to represent all words and sentences in Mini-Basque using Adobe Photoshop. Examples of images are shown in Figure 2 and Figure 3. In the image pilot, each individual animal and color was presented once. Additionally, subjects saw thirty-two scenes depicting a representative sample of sentence-level stimuli (i.e., each animal appeared as the subject and object of each verb, in both singular and plural forms). Throughout the 40 total items, each animal appeared in each color at least once. Five native speakers of English were asked to describe the 40 images. Their responses suggested that the images did indeed represent their intended concepts. Two native speakers of
Basque (one male, one female) saw the same images were asked to judge whether the words and sentences in Mini-Basque appropriately described the associated image. They both accepted 100% percent of the word-picture and sentence-picture pairs.

The image pilot task differed for Basque and English speakers because of expectations for how they would approach a picture-matching task (see Section 2.2.3.1). Native English speakers should hear a Basque word (or sentence) and map it to a concept associated with the correct picture, which should match the actual concept associated with the Basque word. Native Basque speakers, however, should already have a concept for the Basque word, and should simply need to decide which of two pictures appropriately represents that concept.

2.2.1.3 Features of Mini-Basque. In this section, I describe the characteristics of Mini-Basque. Mini-Basque is a subset of the full Basque language, so all of its features do exist in the natural language. Features described as pertaining to “Basque” refer to general features of Basque that are present in the mini-language, whereas features described as pertaining to “Mini-Basque” refer to specific characteristics that are true for all Mini-Basque sentences, but for only a subset of the actual Basque language.

The final version of Mini-Basque contained 12 concrete content (open-class) words, all of which are frequent in both Basque and their English translations: four nouns (animals), four verbs (transitive; see Section 2.2.1.2), and four adjectives (colors); see Figure 2. None of these are cognates with English or Romance languages. Additionally, Mini-Basque contains one function (closed-class) word (ukan: to have), which is a frequent auxiliary verb in Basque. This word is learned only in the sentence context, as a part of Basque grammar.
<table>
<thead>
<tr>
<th>Nouns</th>
<th>Behia “cow”</th>
<th>Hartz “bear”</th>
<th>Txakurra “dog”</th>
<th>Zaldia “horse”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>![Cow Image]</td>
<td>![Bear Image]</td>
<td>![Dog Image]</td>
<td>![Horse Image]</td>
</tr>
<tr>
<td>Verbs</td>
<td>Bultzatu “to push”</td>
<td>Ikusi “to see”</td>
<td>Miazkatu “to lick”</td>
<td>Usaindu “to smell”</td>
</tr>
<tr>
<td>Adjectives</td>
<td>Beltza “black”</td>
<td>Gorria “red”</td>
<td>Horia “yellow”</td>
<td>Urdina “blue”</td>
</tr>
<tr>
<td></td>
<td>![Black Image]</td>
<td>![Red Image]</td>
<td>![Yellow Image]</td>
<td>![Blue Image]</td>
</tr>
</tbody>
</table>

Figure 2. Sample images and citation forms for the 12 open-class content words and corresponding images in Mini-Basque.

Following the canonical word order of Basque, all sentences in Mini-Basque are SOV. Unlike English, which is a head-first language, Basque is a head-final language, so the determiner follows the noun in a noun phrase (NP). Whereas in English, we would say “the dog,” the Basque word order is literally “dog the” (txakurra; where –a is the determiner). When an adjective is present in a Basque NP, it immediately follows the noun, and the determiner affixes to the adjective. The complexity of Mini-Basque is manipulated by the presence or absence of adjectives on subjects and objects. Figure 3 includes an example of a sentence that does not include an adjective on the subject, but does on the object.

Basque morphology is complex, as determiners carry information about both number and case marking. The following formal descriptions of Basque morphology and morphosyntax are based on the grammars of de Rijk (2008) and Laka (1996). The singular determiner in Basque is...
-a, which is a bound morpheme affixed to the end of the noun or adjective. The morpheme -k indicates plurality, so the plural form of the determiner becomes -ak. However, Basque is an ergative-absolutive language, which results in the subject of a transitive verb being marked differently for case than the object of a transitive verb (which, in this case, in unmarked). In Basque, the ergative case marking is also realized as -k, so the determiner on the singular subject of a transitive verb becomes -ak. When the plural determiner and the ergative case marking are combined (i.e., ak-+k), this is realized as -ek. Note, therefore, that the -ak ending in Basque is ambiguous, because it can appear on a singular subject or a plural object (see Table 1).

Morphological ambiguity is a common feature in natural languages (e.g., in English, “-s” is used as a plural marker on nouns, a 3rd person singular verb ending, and a possessive marker).

Morpho-syntactically, Basque has polypersonal number agreement, where the verb agrees with both the subject and the object. This can be contrasted to English, in which the verb agrees only with the subject. In Mini-Basque sentences, the auxiliary immediately follows the verb and is inflected to agree with both subject and object number as follows: singular subject and singular object, du; plural subject and singular object, dute; singular subject and plural object, ditu; plural subject and plural object, dituzte (see Table 2).
Table 1. The realizations of the Basque determiner according to case and number.

<table>
<thead>
<tr>
<th></th>
<th>Subject</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td>-ak</td>
<td>-a</td>
</tr>
<tr>
<td><strong>Plural</strong></td>
<td>-ek</td>
<td>-ak</td>
</tr>
</tbody>
</table>

*Note: -ak is ambiguous because it can denote either a singular (ergative) subject or a plural (absolutive) adjective.*

Table 2. The realizations of the Basque auxiliary, *ukan*, according to polypersonal agreement.

<table>
<thead>
<tr>
<th></th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td>du</td>
</tr>
<tr>
<td><strong>Plural</strong></td>
<td>dute</td>
</tr>
</tbody>
</table>

All Mini-Basque sentences are in the present perfective tense, which indicates the recent past (see Figure 3). This tense was chosen because it has relatively simple morphosyntax, in which all verb agreement happens on the auxiliary rather than the main verb, and is therefore the same for all verbs. Given the parameters described in this section, there are 6,400 possible unique sentences in Mini-Basque.


<table>
<thead>
<tr>
<th>Basque</th>
<th>Hartzek zaldi urdina bultzatu dute.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloss</td>
<td>Bear-PL.SUBJ.DET horse blue-SG.OBJ.DET push have</td>
</tr>
<tr>
<td>English</td>
<td>The bears (have) pushed the blue horse.</td>
</tr>
</tbody>
</table>

**Trial example with agreement foil**

(a) correct

(b) agreement foil

**Other foils**

(c) lexical foil

(d) word-order foil

Figure 3. Sample Mini-Basque sentence with associated images. In the picture-matching task, the subject would hear the Basque sentence and see the correct image and one of the three foil images, depending on the condition.

### 2.2.2 Auditory stimuli

A total of 12 word sound files (one per word) and 888 sentence sound files (744 grammatical, 144 ungrammatical) were created for this study. All words (including content word roots, plural and singular forms of nouns and adjectives, and function words) and a subset of sentences were recorded by a female native Basque speaker. The subset of recorded sentences included 80 grammatical sentences, which included two iterations of every NP in both the subject and object position (e.g., *hartz gorriak*; trans.: *the red bear*), and at least two iterations of every verb/auxiliary pair (e.g., for “push”: *bultzatu du, bultzatu dute, bultzatu ditu*, and *bultzatu dituzte*). Thirty-five ungrammatical sentences were also recorded, which
included two iterations of every ungrammatical NP or verb/auxiliary pair present in the list of 144 ungrammatical sentences.

Each word was recorded twice in Audacity 1.3 using a Logitech USB Desktop Microphone. For this experiment, only the citation forms (see Figure 2) were used. Each recording was listened to individually, and the best recording of each word was chosen. The sound files were trimmed to reduce silence in Audacity, and then normalized to similar intensities. Using a Praat script, 100 ms of silence was added to the beginning and end of each word. Including the 200 ms of silence, words ranged from 0.88 to 1.25 seconds ($M = 1.04$, $SD = 0.10$).

Each sentence was also recorded twice and listened to individually, and the best recording of each NP (chosen separately for subjects and objects) and verb/auxiliary pair was extracted, with minimal silence at the beginning and end of each segment. The length of pauses between each phrase in the naturally occurring speech was also measured. The segments were normalized to similar intensity levels in Audacity 1.3. Using a Praat script, these were then spliced together along with pauses whose lengths were determined based on the average pause lengths from a sample of the original recorded sentences. The pause length between the subject NP and the object NP was slightly shortened to reduce the average length of the sentences, and the pause length between the object NP and the verb phrase was slightly lengthened to ensure that these phrases remained distinct in the speech stream. The arrangement of pauses and phrases is shown in Figure 4. A native speaker of Basque listened to a subset of the final sentences and determined that they sounded natural, in spite of the modified pause lengths. After the sentences were spliced together, each new sentence was listened to individually, and each phrase of the
sentence was normalized again in Audacity so that the intensity across phrases within a sentence was similar. Sentences ranged from 2.85 to 5.12 seconds ($M = 4.24, SD = 0.42$).

Auditory stimuli for the control task (see Section 2.2.3.2) consisted of rotated speech, or speech altered so that it is incomprehensible, but has acoustic properties similar to normal speech (Blesser, 1972), of the Basque words and sentences. The 12 word sound files and 12 representative sentence files underwent a low-pass filter of 5000 Hz, and then all remaining non-zero frequencies were rotated around the midpoint (Darwin, n.d.). For each of the 24 rotated files, a “high” version was created by randomly replacing 500ms of the speech stream with an 800 Hz tone, and a “low” version was created by randomly replacing 500 ms of the speech stream with a 250 Hz tone. Thus, a total of 48 sound files were used in the control task – 24 for the word-level task, and 24 for the sentence-level task, half with high tones and half with low tones.

![Figure 4. Praat script arrangement of pauses and phrases in Mini-Basque sentences.](image)

2.2.3 Tasks. This study involved three main tasks that were used to train or test learners on Mini-Basque: a forced-choice picture matching task, a grammatical violation task, and a production task. Training was consistent with implicit training conditions in SLA (DeKeyser, 1995; Norris & Ortega, 2001; Spada & Tomita, 2010). In addition, subjects completed more traditional tasks probing L2 knowledge at the end of training and retention.
2.2.3.1 Picture matching. Subjects were trained by means of a “forced-choice” picture-matching task, in which they heard a word or sentence and chose the picture that best matched what they heard (see Figure 3). Picture-matching tasks, which expose learners to the form-meaning associations that are fundamental to language, are commonly used in language acquisition studies, as well as in foreign language classrooms (Sanz, Lin, Lado, Bowden, & Stafford, 2009; Stafford, 2011; Stafford, Sanz, & Bowden, 2010; VanPatten & Cadierno, 1993). Crucially, picture-matching involves task-essential practice. That is, the learner cannot successfully complete the task without using information about the words and structures being learned (Loschky & Bley-Vroman, 1993). Therefore, this task forces learners to attend to, and hopefully establish, form-meaning connections in Mini-Basque, and thus encourages language learning (VanPatten, 2002). Additionally, studies investigating the neural bases of language have shown that picture matching tasks can elicit differential activation for lexical/semantic and syntactic processing (Manenti, Cappa, Rossini, & Miniussi, 2008; Meltzer, McArdle, Schafer, & Braun, 2010; Skeide, Brauer, & Friederici, 2014; Skeide, Friederici, & Brauer, 2011) and can elucidate learning-related changes in the brain (Schmithorst, Holland, & Plante, 2007).

In the picture-matching task, subjects heard a word or sentence, and then (after a 250 ms fixation cross) saw two pictures. One of the pictures (the “target”) matched the word or sentence, and one (the “foil”) did not. Subjects were instructed to choose the picture that matched what they heard by pressing a button, at which point an asterisk appeared on the screen. After a fixed amount of time (2000 ms for words and 3500 ms for sentences, based on pilot testing), the incorrect picture disappeared, and the correct picture remained on the screen for 1000 ms, providing feedback intended to promote learning (Potowski, Jegerski, & Morgan-Short, 2009;
Note that the timing for this task in and out of the MRI scanner was exactly the same, except for the inter-trial interval (ITI). When performed outside of the scanner (for the word-level only), the ITI for this task was the 500 ms fixation cross at the beginning of each trial. When performed in the scanner, feedback was followed by a fixation dot of variable duration ITI (0 to 12.5 seconds), and this fixation continued until the next pulse sequence, which happened every 2.5 seconds and was linked to the repetition time (TR) of the scan (see Section 2.6.1). This was then followed by a 500 ms fixation cross, which alerted the participant to the upcoming trial. Details of the time course of the word- and sentence-level versions of this task are shown in Figure 5 and Figure 6, respectively.

In the word-level picture-matching task, the target and foil always belonged to the same word class (i.e., noun, verb, or adjective). They were minimal pairs, differing along only one dimension. For example, for the word *hartza* (trans: *bear*), the subject might see a picture of a bear and a picture of a cow, but both would be the same color. For the word *urdina* (trans: *blue*), the subject might see a picture of a blue dog and a picture of a red dog, but never two different animals. For the word *bultzatu* (trans: *push*), the subject might see a picture of a yellow horse pushing blue bears and a picture of a yellow horse smelling blue bears. That is, for verbs, the animals, colors, and numbers were always the same for both the agent and patient in both pictures; only the action was different. A total of 13 word-level picture-matching runs were created for the training and retention sessions of this experiment. Each run contained 48 trials, with four exemplars per word. Over the course of the experiment, there were a total of 624 word-
picture matching trials, and 52 exemplars per word. Target and foil images were distributed evenly within and across runs according to animal, color, and number.

Figure 5. Timing of word-level picture-matching task (objects not drawn to scale). When performed outside of the scanner, the last two events were not included.

In the sentence-level picture-matching task, targets and foils were also minimal pairs. In this case, the target picture represented the sentence, and the foil only differed from the sentence along one dimension, either lexicon, word order, or agreement. Specifically, the foil either depicted a noun whose number did not match that of sentence, either for the subject or object
(Figure 3a), had an incorrect lexical item (Figure 3c), or showed the agent and patient of the sentence in reversed roles (Figure 3d). The purpose of these three conditions was to assess lexical/semantic, syntactic, and morphosyntactic learning separately. A total of ten sentence-level picture-matching runs were created for the training and retention sessions of this experiment. Each run contained 48 trials, with 16 trials per condition (i.e., lexical, word order, and agreement). Over the course of the experiment, there were a total of 480 unique sentence-picture matching trials, and 160 trials per condition.

The location of targets (left/right) was pseudorandomized and counterbalanced. Two versions of each task were created so that the placement of targets and foils in one version was the opposite of their placement in the other.
Within the field of SLA, training conditions are often characterized as either *implicit* or *explicit*. A training condition is traditionally considered *explicit* if it provides rule explanation or metalinguistic feedback, or if learners are asked to attend to linguistic forms and extract rules on their own (DeKeyser, 1995; Norris & Ortega, 2001; Spada & Tomita, 2010). In this way, explicit training conditions can be either deductive or inductive. *Implicit* training conditions, on the other hand, are characterized by the lack of rule explanation, metalinguistic feedback, and direction to attend to specific forms (DeKeyser, 1995; Norris & Ortega, 2001; Spada & Tomita, 2010). These

Figure 6. Timing of sentence-level picture-matching task (objects not drawn to scale).
types of conditions might include large amounts of L2 input, otherwise known as an input flood, or feedback that is considered to be more implicit (Spada & Tomita, 2010).

The training condition in this study (i.e., the picture-matching paradigm) is consistent with traditional implicit training conditions in SLA. There was no rule explanation, and while participants did receive feedback, it did not contain any metalinguistic information or explanations as to why a particular picture was correct. It simply reinforced the correct answer. This type of feedback may play a similar role to recasts in conversational interaction, in which the interlocutor simply reformulates an incorrect utterance in a more target-like way, without providing any explicit information about the accuracy of the initial utterance or the form of the correct utterance. This type of feedback is considered to be implicit in nature (R. Ellis, Loewen, & Erlam, 2006; Long, Inagaki, & Ortega, 1998; Lyster, 1998; Mackey & Philp, 1998). In lieu of rule explanation, training on Mini-Basque focused on providing learners with a lot of exposure to form-meaning connections in the L2, as is characteristic of implicit conditions in SLA (Spada & Tomita, 2010).

It is important to note, however, that an implicit training condition does not necessarily result in implicit learning, or the development of implicit knowledge, just as an explicit training condition does not necessarily result in explicit learning or explicit knowledge (DeKeyser, 2003; Hulstijn, 2005). It is often the case that both types of learning and knowledge result from either type of training, which can be ascertained with tasks like think-aloud protocols (Leow, 1999;)

---

3 While the Mini-Basque training paradigm is consistent with implicit training conditions in SLA, it would not be considered an implicit condition in cognitive psychology. In psychology, implicit training conditions (e.g., for artificial grammar or serial reaction time tasks) are classically defined as conditions in which participants are not informed about (and often distracted from) the true purpose of the experiment, and learning is therefore assumed to be incidental (Reber, 1976; Reber, Walkenfeld, & Hernstadt, 1991; Willingham & Goedert-Eschmann, 1999). Learners in the Mini-Basque study were always aware that they were learning a language, and it can therefore be assumed that at least some of the learning in this study was intentional.
Leow & Morgan-Short, 2004; Sachs & Polio, 2007; Sanz et al., 2009), debriefing questionnaires or interviews (Hamrick & Rebuschat, in press; Rebuschat et al., in press; Tagarelli et al., 2011, 2014), and online judgments like confidence ratings and source attributions (Dienes, 2004; Dienes & Scott, 2005; Hamrick & Rebuschat, in press; Rebuschat et al., in press; Tagarelli et al., 2014). As discussed below, the current study employs think-aloud protocols and debriefing questionnaires, which can detect explicit processes if a participant demonstrates the ability to verbalize or explain rules. However, if a participant does not verbalize or explain rules but still exhibits evidence of learning, this could indicate that knowledge is implicit, or it may just be due to underreporting or the lack of appropriate measurements to assess explicit knowledge. These are important considerations to keep in mind when interpreting results within the framework of implicit and explicit training, learning, and knowledge in SLA.

2.2.3.2 Control (tone-matching) task. A sensorimotor control task (see Figure 7) was designed to isolate language-related activation by subtracting out task-level activation in the picture-matching tasks, such as auditory, visual, and motor processing, decision making, and attention. It differs from the experimental tasks in two key ways. First, instead of hearing a word or sentence in Basque, participants heard rotated speech, interspersed with either a low or high tone (see section 2.2.2). As described above, the length of the rotated speech was matched to the average length of words or sentences, depending on the task (word or sentence level). Second, participants were presented with two images from the same set as in the experimental tasks. However, both pictures were the same (e.g., at the word level, a subject might see a blue dog on both the left and right sides of the screen). The one difference between the pictures was placement on the screen – one picture was offset 25% above the center of the screen, and the
other was offset 25% below the center of the screen. Subjects were asked to choose the higher picture when presented with a high tone, and the lower picture when presented with a low tone. This is meant to control for the attentional and memory resources involved in hearing a sentence, waiting, and then choosing a picture based on the aural input. The association between a high and low tone and up and down, respectively, is thought to be automatic and therefore should still allow for the isolation of any effects specific to language learning in the experimental task. The number of control items was one-fourth of the number of experimental items, following previous fMRI research (Caplan et al., 2002; Dale, 1999; Manoach, Greve, Lindgren, & Dale, 2003).
Figure 7. Timing of control task (objects not drawn to scale). Word-level is shown. Timing for the sentence-level control task is exactly the same as the sentence-level picture-matching task.

2.2.3.3 *Timed grammaticality judgment task.* Violation paradigms are extremely common in testing L2 knowledge with behavioral and ERP measures, and they have been used in fMRI as well (Bowden et al., 2013; R. Ellis, 2005; Friederici et al., 2002; Morgan-Short, Finger, et al., 2012; Morgan-Short et al., 2010; Morgan-Short, Steinhauer, et al., 2012; Mueller, 2005; Newman et al., 2001a; Opitz & Friederici, 2003, 2004; Rebuschat & Williams, 2006). The violation task in this study is a classic timed GJT. Subjects were presented with sentences, half of
which followed the (morpho)syntax of Mini-Basque, and half of which violated either word
order or agreement rules. A grammatical sentence is shown in example (1). The examples of
ungrammatical sentences below are based on violations of (1).

(1) *Hartz beltzek behi gorria usaindu dute.*
Bear black-DET<sub>pl</sub> cow red-DET<sub>sg</sub> smelled has-PL<sub>sub-SG<sub>obj</sub></sub>
“The black bears (have) smelled the red cow.”

*Syntactic word-order violations* were created by moving an element of the sentence to an
incorrect place. This was done either within a phrase, as show in example (2), or across phrase
boundaries, as shown in example (3). An equal number of violations of each type were applied to
the subject and object NPs.

(2) *Hartzek* beltz-Ø* behi gorria usaindu dute.
Bear-*DET<sub>pl</sub> black-*Ø cow red-the smelled has-PL<sub>sub-SG<sub>obj</sub></sub>

(3) *Hartz* *dute* beltzek behi gorria usaindu *Ø.
Bear *has-PL<sub>sub-SG<sub>obj</sub></sub> black-DET<sub>pl</sub> cow red-the smelled *Ø

*Morphosyntactic agreement violations* were created by changing the auxiliary at the end of the
sentence so that it no longer agreed with the determiners on the subject and object of the
sentence, as shown in example (4). The auxiliary was always changed so that the number for
only one noun would be incorrect. For example, if the correct auxiliary was *du* (SG<sub>sub-SG<sub>obj</sub></sub>), it
was never changed to *dituzte* (PL<sub>sub-PL<sub>obj</sub></sub>).

(4) *Hartz beltzek behi gorria usaindu *du.*
Bear black-DET<sub>pl</sub> cow red-DET<sub>sg</sub> smelled has-*SG<sub>sub-SG<sub>obj</sub></sub>
In each trial of the GJT, participants heard a sentence and, after a 250 ms delay, were asked to judge, by button press, whether the sentence was “good” or “bad” according to what they learned in the picture-matching tasks. For half of the subjects in this study, “GOOD” appeared on the left side of the screen, and “BAD” appeared on the right. This was reversed for the other half of the subjects. Subjects were given 2500 ms to respond, then the task advanced to the next trial, following a variable duration ITI and the onset of the subsequent TR. Like the picture matching tasks, an asterisk appeared when the subject responded. A total of six GJT runs were created for the training and retention sessions of this experiment. Each run contained 48 trials, with 12 trials per condition (i.e., correct syntax, syntactic violation, correct morphosyntax, and morphosyntactic violation). Over the course of the experiment, there were a total of 288 grammaticality judgment trials, and 72 trials per condition. See Figure 8 for details on the timing of the GJT.
To optimize fMRI statistical analyses, the order of trials and the length of ITIs in the picture-matching and violation tasks was determined by OptSeq2 (Dale, Greve, & Burock, 1999). To maximize the number of learning trials while maintaining statistical efficiency for fMRI analyses, the ratio of total trial time to total ITI time was 4:1 (Dale, 1999). For the picture-matching and GJT, the 48 trials in each run were divided into four “mini-blocks” of 12 trials to allow for more fine-grained analyses over time. For the word-level picture-matching task, each mini-block contained one of each of the 12 words, and foils were balanced across mini-blocks. For the sentence-level picture-matching task, each mini-block contained four trials per condition (lexical, agreement, and word order), and for each condition, there was one trial per verb tense (du, dute, ditu, dituzte) in each mini-block. For the GJT, each mini-block contained three trials.
per condition (syntactic violation, morphosyntactic violation, matched syntactically correct, matched morphosyntactically correct), as well as four trials per verb tense. The condition-tense pairs were balanced across mini-blocks.

2.2.3.4 Production. The production task was a simple picture-naming task at the word level and a scene description task at the sentence level. In word-level production, subjects saw a picture depicting a noun, verb, or adjective, and were asked to name what they saw. For example, if the subject saw the picture in Figure 9a, he/she should say *txakurra* (trans.: *dog*) out loud. In sentence-level production, subjects saw a picture depicting a scene that could be described using the vocabulary and grammar they previously learned. They were asked to produce a sentence describing the picture, and to say as much about the picture as they could. For example, if the subject saw the picture in Figure 9b, below, he/she should say *Hartza txakurra miazkatu du* (trans.: *The bear licked the dog*). No feedback was provided during production. Previous research indicates that learners are able to do such tasks after a short training period (Grey, 2013; Hudson Kam & Newport, 2005, 2009; Morgan-Short, 2007). A total of four productions runs for each level (word and sentence) were created for the training and retention sessions of this experiment. Each run contained 24 trials. For words, each word was represented twice per run. For sentences, each plural/singular-subject/object combination was presented six times, and other features of the sentences (animals, colors, verbs) were balanced. Over the course of the experiment, there were a total of 96 word and 96 sentence production trials, with a total of 8 trials per word and 24 trials per number combination (verb tense). Sample answer sheets used in the word- and sentence-level production tasks are shown in Appendix A and Appendix B, respectively.
Figure 9. Word and sentence production. Shown are examples of what a participant might have seen on the screen during the (a) word- and (b) sentence-level versions of the production task. In (a), the subject should say *txakurra*, “dog”. In (b), the subject should say *Hartza txakurra miazkatu du*, “The bear (has) licked the dog.”

2.2.3.5 Additional assessments. Several additional behavioral assessments were conducted to directly tap into aspects of L2 acquisition not covered by the tasks above, including aspects of explicit knowledge and awareness, as well as communicative abilities in less constrained settings. These measures were adapted from previous research in the Brain and Language Lab at Georgetown University (Grey, 2013; Morgan-Short, 2007).

2.2.3.5.1 Written grammaticality judgment task. Subjects completed an untimed written GJT at the end of training and at the end of the retention session. This task contained 48 new sentences: 24 were correct and 24 contained grammatical violations. Half of the violations contained syntactic word-order mistakes and half contained morphosyntactic agreement mistakes (see section 2.2.3.3). Because subjects had never seen the Mini-Basque words written before, they were given a key showing the spellings for each of the 12 content words that they had
learned to use as an aide during this task. In the written GJT, subjects were asked to read the sentence and circle whether it was good or bad. If they chose “bad,” they were asked to correct the sentence and provide an explanation for why the sentence was bad. Subjects were also instructed to “think aloud” while completing the task, which probed their awareness of their own grammatical knowledge during online task completion. Four versions of this task were created. All versions had the same 48 sentences in different randomized orders. Subjects were randomly assigned versions for training, and were pseudorandomly assigned a different version during the retention session. Version 1 of the written GJT is provided in Appendix C.

2.2.3.5.2 Free response task. The free response task was designed to be a less controlled version of the sentence production task. It is similar to the card game, Go Fish. In this task, the subject was presented with five 2x2 grids, each of which contained three pictures of scenes from the mini-language and one blank box. The researcher had five cards depicting the missing scenes from the blank boxes. The subject was instructed to decide which picture was missing from each grid and ask the researcher for that picture by saying the Mini-Basque sentence that described the picture. This task required meaningful communication because the subject and the researcher could not see each other’s materials, so the subject had to adequately communicate to the researcher that he/she was looking for a particular card. An example of a grid and the missing picture is shown in Figure 10. For the production tasks, the written GJT, and the free response tasks, participant responses were recorded with an Olympus VN-702PC digital voice recorder.
2.2.3.5.3 Debriefing questionnaire. After completing all tasks, subjects answered two questionnaires. The first was a debriefing questionnaire (see Appendix D), which was adapted from Morgan-Short (2007), and was designed to assess the extent to which subjects had acquired metalinguistic knowledge and awareness. In this questionnaire, subjects were asked if they noticed or searched for rules in Mini-Basque, and whether they could state any rules. The questionnaire was designed to allow subjects to first state any general rules that they noticed, and then got more specific in order to probe knowledge about nouns, adjectives, verbs, determiners, and word order.

2.2.3.5.4 Wrap-up questionnaire. Finally, subjects completed a very brief questionnaire asking them whether they had tried to figure out what language they were learning over the course of the study, and whether they had a guess as to which language it was. This served to
ensure that no subjects had any outside exposure to Basque over the course of the study. The wrap-up questionnaire (see Appendix E) also included a question asking subjects to briefly describe any language classes in which they were currently enrolled, which could potentially be included as a covariate in analyses. No subjects reported trying to figure out the language, and none knew that the language was Basque.

2.2.3.6 Demographic information. All subjects completed the standard “Info General” background questionnaire that is administered to all subjects who participate in studies in the Brain and Language Lab, as well as the Edinburg Handedness Inventory (Oldfield, 1971). They were also screened using the “Safety Screening Form for Magnetic Resonance (MR) Procedures” that is administered to all participants at the Center for Functional and Molecular Imaging at Georgetown University.

2.3 Pilot testing

Four small-scale behavioral pilot studies were conducted to test the materials, ensure that Mini-Basque was learnable in a short period of time, and determine the most effective procedure for the order and number of word-level, sentence-level, and timed GJT runs. The materials used in the pilot studies are the same as those described above, unless otherwise stated.

2.3.1 Behavioral Pilot 1. Nine native speakers of English (6 females) participated in the first behavioral pilot, which aimed to assess the learnability of Mini-Basque. One participant was excluded because of a technical problem during the experiment. Subjects completed two runs of 96 word-picture matching trials each (16 per word), followed by two runs of 150 sentence-picture matching trials each. They then completed short word and sentence production tasks. The debriefing questionnaire was also piloted. Subjects reached, on average, 98.18% accuracy on
words and 79.75% accuracy on sentences. Analyses revealed significant improvement over time in both accuracy and RT, and different patterns for lexical, agreement, and word-order trials. Importantly, this pilot revealed evidence of mini-language learning over a short period of time (about one hour), and that language domains could be distinguished, at least behaviorally.

2.3.2 Behavioral Pilot 2. Nine native speakers of English (5 females) participated in the second behavioral pilot study, which aimed to optimize the procedure for the study. Subjects participated in three sessions over the course of three days. All word, sentence, and GJT runs consisted of 48 trials, and production runs consisted of 24 trials each for words and sentences. On Day 1, subjects completed three runs of the word-picture matching task, followed by two runs of the sentence-picture matching task, two runs of the GJT, and word and sentence production. On Day 2, they completed one word-picture matching run, followed by three sentence-picture matching runs and word and sentence production. On Day 3, they completed one word-picture matching run, followed by two sentence-picture matching runs, two GJT runs, and word and sentence production. These runs were intended to provide enough training to learn Mini-Basque in three one-hour MRI sessions. However, to test whether additional training would be useful, another set of one word-picture matching run, four sentence-picture matching runs, and one run each of word and sentence production were completed on Day 3. Therefore, there was a total of six word-level picture-matching runs (288 trials, 24 per word), eleven sentence-level picture-matching runs (528 trials, 176 per condition), four GJT runs (192 trials, 48 per condition), and four production runs at the word and sentence level. The control task and all additional L2 tests (written GJT, free response, and debriefing questionnaire) were piloted at this stage.
Subjects reached 97% accuracy on the word-level task, and 75.7% accuracy on the sentence-level task. Again, performance increased over time, with different patterns for each condition. In the GJT, subjects reached 92% accuracy on syntactic violations by Day 3, but only performed at 26% accuracy on morphosyntactic violations, which was actually a decrease in performance from Day 1.

Although subjects in this study had nearly twice as much exposure to the language as those in the first behavioral pilot, they performed slightly worse on the sentence-level picture-matching task. There are several reasons for why this might be the case. First, learners in the first pilot were presented with 192 words and reached ceiling on this task before ever hearing a sentence, whereas learners in the second pilot heard 144 words (four fewer exemplars per word), and were only performing at about 90% accuracy on words before they started learning the grammar. Second, although learners had additional training, this was at the end of a very long session (about three hours), so fatigue effects may have come into play and prevented them from demonstrating further improvements behaviorally, even if they were learning more at a level that was not detectable by behavioral measures. Finally, two different native Basque speakers recorded the stimuli for the first and second pilot. For the first pilot, each sentence was recorded individually, and the rate of speech was very slow, which likely made it easier to understand. For the second pilot, sentences were re-recorded by a different speaker and created as described above in section 2.2.2. These sentences were much more natural and relatively fast, and therefore may have been more difficult for learners to understand.

Another problem with this pilot was the performance on the GJT, which seemed to be related to the violation recordings. For this pilot, full ungrammatical sentences were not recorded
– instead, syntactic violations were created by splicing together pieces of grammatical sentences and incorrect words that had been recorded individually (i.e., not in a sentence context). Morphosyntactic violations, on the other hand, were created in the exact same way as grammatical sentences, but the verb phrase added at the end was incorrect. This resulted in clear acoustic differences between syntactic violations and the other sentence types, which may have caused subjects to group grammatical sentences and morphosyntactic violations together and therefore perform very poorly on morphosyntactic violations.

Adjustments were made to the design prior to the third pilot in order to address these issues.

2.3.3 Behavioral Pilot 3. Prior to the third behavioral pilot, all Basque sentences were re-recorded, including those with syntactic violations, exactly as described in section 2.2.2. These were the final stimuli used in the fMRI experiment. They were slightly slower than those recorded for the second pilot, but still produced at a very natural pace. Importantly, the new recordings did not contain obvious acoustic cues to distinguish syntactic violations from other sentence types.

The main goals of the third behavioral pilot were to 1) test materials with improved auditory stimuli, 2) determine whether it would be possible to decrease the number of word trials and therefore fit more training in the scanner, and 3) determine whether additional performance improvements would be seen if the additional training (i.e., out of the MRI scanner) was moved to the end of Day 2, thereby avoiding fatigue effects on Day 3 but maintaining the amount of exposure from the second pilot.
Four subjects (3 females) participated in the third pilot. The number of trials per run was the same as in the second pilot, but the control task was integrated into any picture-matching runs that would be performed in the MRI scanner in the final experiment. Day 1 consisted of two word-picture matching runs, followed by two sentence-picture matching runs, two GJT runs, and word and sentence production. Day 2 consisted of two word-picture matching runs, three sentence-picture matching runs, and word and sentence production. Additional training was then placed after production, which again included two word-picture matching runs, three sentence-picture matching runs, and word and sentence production. Day 3 consisted of one word-picture matching run, two sentence-picture matching runs, two GJT runs, and word and sentence production. Therefore, there was a total of seven word-level picture-matching runs (336 trials, 28 per word), ten sentence-level picture-matching runs (480 trials, 160 per condition), four GJT runs (192 trials, 48 per condition), and four production runs at the word and sentence level. All other materials were also piloted in this study.

Subjects reached 97% accuracy on words by the end of training, and 75% accuracy on sentences. Once again, performance increased over time, with different patterns for different sentence conditions.

The third pilot demonstrated that moving the additional training to the end of Day 2 did not result in improved performance by the end of training. Additionally, presenting learners with only 96 words before they started grammar learning, at which point they were only averaging 83% accuracy on words, was not sufficient for learning grammar to high proficiency.

In the final pilot, any additional training was removed, since it would have made learning in the scanner discontinuous (contrary to one of the main the goals of the current study) and did
not seem to have a positive effect on learning. Since the amount of word training before sentence training seemed to have a substantial effect on grammar learning outcomes, the number of word trials preceding the first sentence run was increased substantially to 240 (20 exemplars per word). As a result, the majority of word learning was planned to occur outside of the MRI scanner in the main fMRI experiment, since changes in grammar learning and the attainment of high proficiency at the grammar level were of greater interest for this study.

**2.3.4 Behavioral Pilot 4.** Four subjects (3 female) participated in the fourth behavioral pilot. One subject did not complete the study. On Day 1, subjects completed five word-picture matching runs, two sentence-picture matching runs, two GJT runs, and word and sentence production. On Day 2, subjects completed two word-picture matching runs, four sentence-picture matching runs, and word and sentence production. On Day 3, subjects completed three word-picture matching runs, two sentence-picture matching runs, two GJT runs, and word and sentence production. Therefore, there was a total of ten word-level picture-matching runs (480 trials, 40 per word), eight sentence-level picture-matching runs (384 trials, 128 per condition), four GJT runs (192 trials, 48 per condition), and three production runs at the word and sentence level. All other materials were also piloted in this study.

Subjects reached 99% accuracy on the word-picture matching task by the fifth run, before starting any sentence-level training. They reached 90% accuracy on the sentence-picture matching task by the end of Day 2. Once again, performance improved over time, with different patterns for different sentence conditions. Significant improvements were seen for both morphosyntactic and syntactic violations in the GJT, but subjects were only clearly able to distinguish syntactic violations (but not morphosyntactic violations) from correct sentences on
either day. The fourth pilot demonstrated that this particular design resulted in high performance compared to previously tested designs, particularly in sentence-level learning. Although subjects still seemed to be having trouble with morphosyntactic agreement, fMRI measures might be able to determine whether learners have some unconscious knowledge of these rules. Therefore, the procedure for the fourth pilot was used in the final fMRI experiment. Details of which runs are included in the MRI scanner are described in Section 2.4 and Figure 11. The fMRI design was tested on two pilot subjects, who were included in the final fMRI analyses and are therefore discussed in the main analyses.

2.4 Procedure

The entire study consisted of five sessions on separate days, lasting from about 2.5 to 3.5 hours each. The entire procedure for Mini-Basque training and testing is outlined in Figure 11, and each session is described in detail below. Prior to scheduling their participation, all potential subjects underwent an extensive phone interview to ensure that they qualified for the study based on background information and MR safety parameters. Following standard procedures at the Center for Functional and Molecular Imaging (CFMI) at Georgetown University, all subjects underwent the MR safety screening procedures twice on each day that involved scanning, and all female participants took a pregnancy test.
Figure 11. Detailed fMRI procedure. A) Order of learning and testing tasks on each day of training. The light grey box indicates tasks performed inside the MRI scanner. B) Sample timing of trial onsets in the sentence-picture matching task. C) Sample timing of trial onsets in the GJT.
2.4.1 Day Ø: Screening and cognitive testing. During the first session (Day Ø), subjects provided informed consent and completed the “Info General” background questionnaire and the Edinburgh Handedness Inventory (Oldfield, 1971). They then completed a battery of cognitive tests, which will be used in later analyses on the relationship between individual differences and L2 acquisition, but will not be discussed here. This session did not involve any MRI scanning and took about 2.5 hours.

2.4.2 Day 1: Training session 1. The first language training session was scheduled about a week after Day Ø (range = 6 to 7 days, $M = 6.86$ days, $SD = 0.36$ days). In this session, subjects first completed the Modern Language Aptitude Test (Carroll & Sapon, 1959), which, like the other cognitive tests, is beyond the scope of this dissertation. Subjects then completed a short practice exercise in which they learned how to do the control, word-picture matching, and sentence-picture matching tasks with English stimuli. This consisted of four practice control items, as well as three example items, three sample items, and 12 practice items for both words and sentences, evenly distributed across word class and condition. Subjects were given the opportunity to ask questions, and were informed that there would be a GJT and a production task at the end of the session. Practice lasted about ten minutes. After practice, subjects completed four word-level picture-matching runs, which took about four minutes each.

In the scanner, subjects completed one more word-level picture-matching run (~7.5 minutes), two sentence-level picture-matching runs (~14 minutes), and two GJT runs (~9.5 minutes). Afterwards, they completed the word and sentence production tasks outside of the scanner, which took about 15 minutes. The entire session lasted about 3.5 hours.
2.4.3 Day 2: Training session 2. The second language training session was scheduled about two days after Day 1 (range = 1 to 4 days, $M = 1.86$ days, $SD = 1.10$ days). In this session, subjects first completed follow-up aptitude tests, which will not be discussed here. After this, they completed a short practice task with the same practice items as in Day 1, but without example and sample items (~5 minutes). They then completed two word-level picture-matching runs.

In the scanner, subjects completed four sentence-level picture-matching runs. They completed the word and sentence production tasks outside of the scanner. The entire session lasted about 2.75 hours.

2.4.4 Day 3: Training session 3. The third and final language training session was scheduled about two days after Day 2 (range = 1 to 5 days, $M = 2.14$ days, $SD = 1.35$ days). In this session, subjects completed the exact same practice task as in Day 2, followed by two word-level picture-matching runs. In the scanner, they completed one word-level picture-matching run, two sentence-level picture-matching runs, and two GJT runs. Once again, they completed the word and sentence production tasks outside of the scanner. After production, subjects completed the written GJT, the free response task, the debriefing questionnaire, and the wrap-up questionnaire. The entire session lasted about 3 hours.

2.4.5 Retention. Twelve out of 14 subjects returned for a follow-up session about one month after Day 3 (range = 15 to 48 days, $M = 31.45$ days, $SD = 8.91$ days). This session was identical to Day 3, except that the versions of the tasks were different. The entire session lasted about 2 hours.
2.5 Behavioral data analysis

For the picture-matching tasks (including the control task) and timed GJT, accuracy and RT for each trial were automatically recorded by E-Prime. RT distributions are notoriously non-normal, and specifically ex-Gaussian, exhibiting skewness toward the left, with a long positive tail on the right (Lachaud & Renaud, 2011; Luce, 1986; Whelan, 2008). Because standard statistical measures, such as ANOVAs, assume a normal distribution, RT data must be treated to be more appropriately suited for these tests. To that end, all RT data was treated as follows. First, RT analyses were performed on correct responses only. RTs faster than 200 ms were discarded, because it was assumed that these responses were too fast to represent the processing involved in a decision and motor response. Because all tasks imposed a relatively short time limit, it was not necessary to remove very long RTs that might otherwise be due to inattention. Next, a natural logarithmic transformation was applied to all RT data points. Outliers (± 3 SD from the mean) were excluded based on the log-transformed data by subject (for each run) and item. RT treatment resulted in the loss of 0.7% of data points for correct responses. Subject-level means were calculated for each subject from the remaining filtered, log-transformed data, and these were entered into repeated-measures ANOVAs. Note that for the GJT, the same steps were followed, but on both correct and incorrect trials, because analyses included both types of responses.

For the picture-matching tasks, the proportion of correct responses was submitted to an arcsine transformation. This transformation increases normality of otherwise skewed categorical data (Cohen & Cohen, 1983; Dixon, 2008), which was the case here, particularly at the word
level and in later sentence runs, when the proportion of correct responses approached 1.0. These transformed data were entered into repeated-measures ANOVAs.

For the timed GJT, grammaticality judgment scores were transformed to $d'$ scores, which are sensitive to response bias and therefore provide a more appropriate measure of an individual’s ability to discriminate between trial types (i.e., grammatical v. ungrammatical) than raw accuracy measures (Macmillan & Creelman, 2005; Stanislaw & Todorov, 1999; Wickens, 2002). $D'$ scores were calculated for each type of sentence (syntactic violation, correct syntax, morphosyntactic violation, correct morphosyntax) for Day 1, Day 3, and Retention (the two runs from each day were combined to increase power), using the formula in (5). $D'$ scores were entered into repeated-measures ANOVAs. While analyses were performed on transformed accuracy and RT data, untransformed data are also reported below for ease of interpretation.

(5) $d' = z(\text{hit rate}) - z(\text{false alarm rate})$

For the production tasks, RT (i.e., voice onset time) was automatically recorded by E-Prime, and participant responses were digitally recorded. RT analyses for production will not be discussed here. For the word production task, the researcher indicated whether the response was correct or not during the task. When the response was not correct, the researcher transcribed the subject’s response. Another researcher later listened to and transcribed all responses for the word production task using SAMPA (Speech Assessment Methods Phonetic Alphabet), a computer-readable version of the International Phonetic Alphabet (Wells, 1997). A word was counted as incorrect if there were more than two phonemes that did not share at least three distinctive features with the correct phoneme (e.g., if the subject said /gorreja/ instead of /gorrija/, this is still correct because both /e/ and /i/ are front, non-low vowels).
For the sentence production task, the researcher indicated whether each element of the sentence was correct, transcribed any incorrect words in the sentence, and noted any word order mistakes. Sentence production was coded as follows: for each sentence, subjects received one point for using SOV word order, one point per NP for using noun-adjective order, one point per NP for using a determiner, one point per NP for using the correct determiner, one point for using an auxiliary verb, one point for using the correct auxiliary verb, and one point per correct lexical item. Because the focus of the sentence production task was on grammar, coding of lexical items was more lenient than in the word production task. For the most part, coding of sentence production involved assigning points based on the provision of a form in an obligatory context. However, a very common mistake in sentence production involved duplicating the determiner between the noun and adjective and again after the adjective (e.g., *txakurrak* gorriak instead of *txakur gorriak*). In this case, the form was provided in an obligatory context, but also provided in an incorrect context. Subjects therefore received one point when they did not include a determiner between the noun and adjective. Subjects could receive a maximum of 14-16 points per sentence, depending on the number of adjectives in the sentence. The coding system allows for separate analyses of word order, morphosyntax, and lexical knowledge, but only results from overall scores will be discussed below.

Results from the written GJT, free response task, and questionnaires are not discussed here.

2.6 fMRI Acquisition and Analyses

2.6.1 FMRI data acquisition. FMRI data were acquired via continuous scanning on a Siemens 3T Trio whole-body MR system with a 12-channel head coil at the CFMI at
Georgetown University, using a T2*-weighted gradient-echo EPI sequence (TE = 30 ms; TR = 2.5s; flip angle = 90°; matrix size = 64x64 mm; field of view = 192 mm; voxel size = 3.0 mm³). Forty-seven axial slices are acquired in interleaved order, with a slice thickness of 3mm (no gaps between slices), covering the full cerebrum and the full cerebellum. One high-resolution (1x1x1 mm³) 3D T1-weighed anatomical scan (MP-RAGE) was acquired for each subject.

Stimuli were presented with E-Prime 2.0 in an event-related design. Behavioral (button-press) responses were collected on two fiber optic response boxes (Psychology Software Tools, Inc.). At the beginning of each scanning session, a structural localizer was run (TE = 2.53; TR = 5.05; flip angle = 58°; field of view = 256 mm; 30 slices; voxel size = 2.0 x 2.0 x 4.0mm) to determine the location of the brain during each session. For each run, three dummy scans (7.5 s) preceded the functional scanning to ensure tissue steady-state magnetization. These dummy scans were discarded. Each run began with an instruction screen and advanced to a 10 s (4 TR) fixation cross after the first pulse sequence to allow time for T1 equilibration effects. For the GJT runs, this fixation was followed by the 48 trials. For the picture-matching runs, this fixation was followed by six control trials, then a 7.5 s (3 TR) instruction screen signaling a switch to the word or sentence task, then the 48 trials. This was then followed by another 7.5 s instruction screen signaling a switch to the control task, and six more control trials. The control trials were placed at the beginning and end of the experimental task to account for signal drift (Smith et al., 1999). Each run ended with a 20 s (8 TR) fixation cross to allow time for T1 equilibration effects at the end of the scan. For details of fMRI task timing, see Figure 11.

Auditory stimuli were presented through Sensimetrics S14 Insert Earphones (Hearing Components Inc.) with Comply Canal Tips, and protective earmuffs (Bilsom Thunder T2) were
worn, for a total noise reduction of at least 37 dB (Berger, 1983, 1984). Additionally, the Tim (Total Image Matrix) upgrade offers shorter acquisition times and/or significant improvement in the signal-to-noise ratio, as well as a 90% reduction in sound pressure. This, combined with the earphones and ear defenders, provides sufficient noise reduction for continuous fMRI scanning during auditory language input (Arnaud, Sato, Ménard, & Gracco, 2013; Powers, Hevey, & Wallace, 2012; Steinmann & Gutschalk, 2012; Venezia, Saberi, Chubb, & Hickok, 2012). A mirror mounted on the head coil allowed subjects to see images, which were back-projected on a translucent screen located at the rear of the scanner.

2.6.2 MR data analysis. FMRI analyses were completed using the SPM8 software package (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) for MATLAB (The Mathworks, MA). Before pre-processing, all EPI images were manually reoriented to align with the SPM T1-weighted template image. Pre-processing included slice timing correction, which temporally corrected EPI images to the middle slice, spatial realignment for motion correction, coregistration of EPI scans to structural scans, segmentation of MP-RAGE volume by tissue type, spatial normalization to a standard MNI reference brain with voxel size of 2mm\(^3\), and smoothing with a Gaussian kernel of 6mm full width half-maximum (FWHM).

The first-level analysis used a fixed-effects single-subject analysis convolving trial type onset vectors with the canonical hemodynamic response function. Item-level variables (e.g., sentence conditions) were introduced into the model at this level. Contrasts from the first-level analyses were smoothed with a Gaussian kernel of 4mm FWHM and entered into second-level random-effects group analyses, at which point subject-level continuous variables, such as accuracy, were included in the model.
For the word- and sentence-level picture-matching tasks, the main analyses consisted of contrasts between the experimental task, which was modeled on the onset and duration of each word or sentence in the task, and the control task (baseline), which was modeled on the onset and duration of each rotated speech event. The feedback event for each trial type was also modeled, this time as the onset of the presentation of the correct response with a duration of 1000 ms. Separate conditions (i.e., noun, adjective, verb and lexical, agreement, word order) were modeled as separate events, but these events were combined in most analyses in order to get a sense of general brain patterns during language learning. For the picture matching tasks, the following contrasts were run at the whole-brain level for the aural presentation of the word or sentence (in both directions, where appropriate):

1) Word or Sentence v. Baseline, collapsed across all runs

2) Word or Sentence v. Baseline, Day 1 (Runs 1 and 2 at the sentence level)

3) Word or Sentence v. Baseline, Day 3 (Runs 7 and 8 at the sentence level)

4) [Word or Sentence v. Baseline at Day 1] v. [Word or Sentence v. Baseline at Day 3]

5) [Word or Sentence v. Baseline at Day 3] v. [Word or Sentence v. Control at Day 1]

6) [Word or Sentence v. Fixation at Day 1] v. [Word or Sentence v. Fixation at Day 3]

7) [Word or Sentence v. Fixation at Day 3] v. [Word or Sentence v. Fixation at Day 1]

Contrasts 4 through 7 were only run in one direction because 4 and 5, as well as 6 and 7, are inverse contrasts (e.g., [Day 1 Word – Day 1 Baseline] – [Day 3 Word – Day 3 Baseline] = [Day 3 Baseline – Day 3 Word] – [Day 1 Baseline – Day 1 Word]). Contrasts 6 and 7 were run against fixation because any task effects should be subtracted out between Day 1 and Day 3, so highly controlled baseline was less important for these contrasts. All of the above contrasts were also
run for the feedback event at the sentence-level, because this event more clearly distinguishes the	hree sentence conditions. Additionally, contrasts of the linear increase and decrease across all
eight sentence runs against baseline and fixation were also conducted, in which each subsequent run either increased or decreased in weight in a stepwise fashion, from -7 to 7.

Contrast 1 on the feedback event was performed separately for the three sentence-level conditions (lexical, agreement, and word order). Because the timing between the feedback event and previous events is constant, it is impossible to fully deconvolve the hemodynamic response function (HRF) from this event. However, by averaging across trials, the influence from the HRF from the previous events should be the same across conditions, so any differences in activation between conditions should be due to differences in the feedback event according to condition.

For the timed GJT, the experimental task was again modeled as the onset and duration of each sentence. The grammatical (GR) sentences served as the baseline items for this task. The following contrasts were run at the whole-brain level for both syntactic and morphosyntactic violations (UG), in both directions:

1) UG v. GR, collapsed across all runs
2) UG v. GR, Day 1 (Runs 1 and 2)
3) UG v. GR, Day 3 (Runs 3 and 4)
4) [UG v. GR at Day 1] v. [UG v. GR at Day 3]
5) [UG v. GR at Day 3] v. [UG v. GR at Day 1]

The statistical threshold was set to $p < .05$, with a family-wise error (FWE) correction for multiple comparisons.
2.6.2.1 Region of interest analyses. Regions of interest (ROIs) were extracted in SPM8 using the anatomical area label (AAL) templates in MarsBaR (http://marsbar.sourceforge.net/). Sixteen ROIs were chosen on the basis of predictions from the DP model and the Convergence Hypothesis, which included the following regions bilaterally: caudate, putamen, hippocampus, parahippocampal gyrus, BA6 (precentral gyrus), BA44 (pars opercularis), BA45 (pars triangularis), and BA47 (pars orbitalis). Beta weights for each ROI, which is an indication of percent signal change, were extracted based on the 2nd-level t-contrasts of task minus baseline for the word, sentence, and grammaticality judgment tasks collapsed across all runs, and for Day 1 and Day 3 separately for words and the GJT. To obtain a more continuous measure of learning, beta weights were also extracted for each of the eight runs of the sentence-level picture-matching task, for both the sentence and feedback events. Syntax and morphosyntax were analyzed separately for the GJT, but all conditions were combined for the word- and sentence-level tasks. One-sample t-tests were conducted between the beta weights of each ROI at each time point and zero to determine whether any ROIs were significantly activated or deactivated. Paired-samples t-tests were conducted to determine whether activation in each ROI changed significantly from Day 1 to Day 3 for the word and grammaticality judgment tasks. For the sentence-level task, one-way ANOVAs were conducted to determine whether there were any changes over time for each ROI. For ROI analyses, \( \alpha \) was set to .05.

2.6.2.2 Analyses probing the relationship between the brain and behavior. Analyses that use the amount of exposure as a predictor of neural activation make the assumptions that subjects learn in a linear fashion, with similar changes in performance in each run, and that all subjects learn in a similar way, with little individual differences between one another. Neither of
these assumptions seems to be true with respect to language learning, so neural activation was also investigated as a function of behavioral performance.

At the word level, the index of behavioral performance for each was the arcsine of accuracy for the very first word-level picture-matching run, which occurred outside of the scanner. This run was chosen because it is the only run that demonstrates clear individual differences between learners, as most learners reached ceiling very quickly. 2nd-level multiple regressions were run in SPM8 on the 1st-level contrast images from the task minus baseline contrasts for Day 1 and Day 3, with the arcsine of accuracy for Run 1 as a covariate. Pearson’s correlations were then run between accuracy for Run 1 and activation in each ROI for Day 1 and Day 3.

At the sentence level, several indices of behavioral performance were tested, including the arcsine of accuracy for Run 1 and Run 8, and the difference between the two (Delta_{8,1}). However, the best performance for each subject was not always in the last run, so Run 8 was not necessarily the best indication of an individual’s potential for acquiring knowledge about the mini-language. Therefore, the arcsine of accuracy for the best run for each subject was also used as a predictor (Run_{MAX}), as well as the difference between Run_{MAX} and Run 1 (Delta_{MAX}). These values were entered as covariates into multiple regressions for activation on Day 1 and Day 3. At the ROI level, the relationship between behavioral performance and activation was tested within-subjects by running Pearson’s correlations between average accuracy and beta weights for each run for all ROIs. It was also tested across subjects by running Pearson’s correlations between accuracy and beta weights, by subject, for each run for all ROIs (e.g., to test whether accuracy in Run 1 was correlated with activation in Run 1 for BA44).
For the timed GJT, several indices of behavioral performance were tested, including $d'$ on Day 1, $d'$ on Day 3, and the difference between the two (Delta$_{3,1}$). Multiple regressions were run with these values as covariates for activation on Day 1 and Day 3. Correlations were then run between Delta$_{3,1}$ and activation in each ROI for Day 1 and Day 3. All analyses were completed separately for syntactic and morphosyntactic violations.

For all reported fMRI results, anatomical locations were determined using MRICron (http://www.mccauslandcenter.sc.edu/micro/mricron/), and BAs were determined using the SPM Anatomy Toolbox (Eickhoff et al., 2005). However, it should be noted that BAs are approximate, as it is impossible to determine the cytoarchitectonics of cortical regions from fMRI data.
Chapter 3: Results

This chapter focuses on results of behavioral performance and fMRI activation from the word- and sentence-level picture matching tasks and the timed GJT. Behavioral results provide evidence of learning, as well as characteristics of the learning process, and the fMRI results reveal patterns of brain activation involved in L2 learning and processing. Results from the word and sentence-level production tasks are also discussed here. Results from additional tasks performed outside of the scanner and the retention session are not reported here.

As noted in Section 2.1, behavioral analyses include all 15 participants. Because of a scanner malfunction for one participant in Day 2 and another participant in Day 3, all analyses for the word-level picture-matching task, the GJT, and sentence-level analyses that exclude Day 2 were conducted on 14 participants (6 female), and any sentence-level analyses including all three scanning days were conducted on 13 participants (5 female).

3.1 Behavioral Results

3.1.1 Word-level learning. All participants completed 10 runs of the word-level picture-matching task over the course of three training days. Average accuracy by run and condition across all subjects is shown in Figure 12 as the proportion of correct responses, and these values, in addition to their arcsine transformations, are reported in Table 3. All statistics were performed on the arcsine transformation of accuracy. A 10 x 3 repeated measures ANOVA with Run and Condition (noun, verb, adjective) as within-subjects factors was conducted. This revealed a main effect of Run, $F(9, 126) = 15.951, p < .001$, partial-$\eta^2 = .533$, observed power = 1.00, a main effect of Condition, $F(2, 28) = 4.261, p = .024$, partial-$\eta^2 = .233$, observed power = .697, and a Run x Condition interaction, $F(18, 252) = 1.666, p = .046$, partial-$\eta^2 = .106$, observed power =
.939. Bonferroni corrected post hoc analyses showed that accuracy on word learning was lower on Run 1 than for all other runs (all ps < .01), with no other differences between runs. Regarding conditions, accuracy was significantly higher on adjectives than verbs. The Run x Condition interaction seems to be driven by increased accuracy on adjectives compared to both nouns and verbs in later runs, where nouns were more accurate than adjectives in Run 2 (p = .053) and there was no difference between verbs and adjectives in early runs (all ps > .2), but adjectives were more accurate than nouns in Runs 6 (p = .025), 7 (p = .018), and 10 (p = .010), and more accurate than verbs in Runs 7 (p = .011), 9 (p = .009), and 10 (p = .017). Additionally, there was a reversal of relative accuracy between adjectives and verbs in Run 8, where verbs were marginally more accurate than adjectives (p = .059). However, none of these differences are statistically significant when corrected for multiple comparisons (Bonferroni corrected α = .00167).

Figure 12. Accuracy on the word-level picture matching task. Accuracy is shown as the proportion of correct responses. Runs highlighted in grey were performed in the MRI scanner.
Table 3. Accuracy on the word-level picture-matching task

<table>
<thead>
<tr>
<th>Run</th>
<th>Adjective Proportion</th>
<th>Adjective Arcsine</th>
<th>Noun Proportion</th>
<th>Noun Arcsine</th>
<th>Verb Proportion</th>
<th>Verb Arcsine</th>
<th>All Proportion</th>
<th>All Arcsine</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.83 (.38)</td>
<td>1.17 (.19)</td>
<td>0.82 (.39)</td>
<td>1.16 (.21)</td>
<td>0.76 (.43)</td>
<td>1.09 (.23)</td>
<td>0.80 (.40)</td>
<td>1.12 (0.13)</td>
</tr>
<tr>
<td>2</td>
<td>0.90 (.31)</td>
<td>1.31 (.21)</td>
<td>0.95 (.21)</td>
<td>1.43 (.17)</td>
<td>0.90 (.30)</td>
<td>1.31 (.21)</td>
<td>0.92 (.28)</td>
<td>1.31 (0.14)</td>
</tr>
<tr>
<td>3</td>
<td>0.95 (.22)</td>
<td>1.41 (.16)</td>
<td>0.95 (.22)</td>
<td>1.41 (.16)</td>
<td>0.94 (.24)</td>
<td>1.37 (.16)</td>
<td>0.95 (.23)</td>
<td>1.38 (0.13)</td>
</tr>
<tr>
<td>4</td>
<td>0.98 (.16)</td>
<td>1.48 (.13)</td>
<td>0.97 (.17)</td>
<td>1.46 (.14)</td>
<td>0.96 (.20)</td>
<td>1.46 (.18)</td>
<td>0.97 (.18)</td>
<td>1.44 (0.12)</td>
</tr>
<tr>
<td>5</td>
<td>0.95 (.22)</td>
<td>1.40 (.15)</td>
<td>0.95 (.22)</td>
<td>1.41 (.16)</td>
<td>0.93 (.26)</td>
<td>1.37 (.19)</td>
<td>0.94 (.23)</td>
<td>1.35 (0.10)</td>
</tr>
<tr>
<td>6</td>
<td>0.99 (.11)</td>
<td>1.52 (.10)</td>
<td>0.98 (.14)</td>
<td>1.49 (.12)</td>
<td>0.93 (.25)</td>
<td>1.38 (.19)</td>
<td>0.97 (.18)</td>
<td>1.42 (0.10)</td>
</tr>
<tr>
<td>7</td>
<td>0.99 (.09)</td>
<td>1.54 (.09)</td>
<td>0.96 (.19)</td>
<td>1.43 (.14)</td>
<td>0.94 (.24)</td>
<td>1.39 (.19)</td>
<td>0.96 (.19)</td>
<td>1.41 (0.10)</td>
</tr>
<tr>
<td>8</td>
<td>0.95 (.21)</td>
<td>1.43 (.17)</td>
<td>0.97 (.17)</td>
<td>1.46 (.14)</td>
<td>0.98 (.13)</td>
<td>1.51 (.12)</td>
<td>0.97 (.17)</td>
<td>1.44 (0.12)</td>
</tr>
<tr>
<td>9</td>
<td>0.99 (.09)</td>
<td>1.54 (.09)</td>
<td>0.98 (.14)</td>
<td>1.50 (.13)</td>
<td>0.97 (.18)</td>
<td>1.44 (.13)</td>
<td>0.98 (.14)</td>
<td>1.46 (0.10)</td>
</tr>
<tr>
<td>10</td>
<td>0.99 (.11)</td>
<td>1.52 (.10)</td>
<td>0.95 (.23)</td>
<td>1.40 (.17)</td>
<td>0.95 (.22)</td>
<td>1.42 (.18)</td>
<td>0.96 (.19)</td>
<td>1.43 (0.14)</td>
</tr>
<tr>
<td>All</td>
<td>0.95 (.22)</td>
<td>1.43 (.04)</td>
<td>0.95 (.22)</td>
<td>1.42 (.03)</td>
<td>0.93 (.26)</td>
<td>1.37 (.03)</td>
<td>0.94 (.23)</td>
<td>1.38 (.10)</td>
</tr>
</tbody>
</table>

Note: Accuracy presented as proportion of correct responses and with an arcsine transformation. Standard deviations are in parentheses.
Average RT by run and condition is shown in Figure 13 as the natural log of RT, and raw RT data and log-transformed RT data are reported in Table 4. All statistics are performed on the log-transformed data. A 10 x 3 repeated measures ANOVA with Run and Condition as within-subjects factors was conducted. This revealed a main effect of Run, $F(9, 126) = 28.569, p < .001$, partial-$\eta^2 = .671$, observed power = 1.00, a main effect of Condition, $F(2, 28) = 70.204, p < .001$, partial-$\eta^2 = .834$, observed power = 1.00, and a marginally significant Run x Condition interaction, $F(18, 252) = 1.628, p = .054$, partial-$\eta^2 = .104$, observed power = .932. Bonferroni corrected post hoc comparisons showed that, as with accuracy, the main effect of Run was driven by poorer performance in very early runs as compared to other runs. Specifically, RT for Run 1 was significantly slower than all following runs ($p < .001$), and RT for Run 2 was significantly slower than Runs 5 through 10 (all $ps < .05$). Regarding condition, there is once again an advantage for adjectives over verbs ($p < .001$), and adjectives were also responded to faster than nouns, but only with marginal significance ($p = .063$). RTs for verbs were also significantly slower than those for nouns ($p < .001$). The marginal Run x Condition interaction seems to be primarily due to the fact that RTs for adjectives and nouns start off almost exactly the same, and RTs are faster for adjectives than nouns in Runs 3 ($p = .020$), 5 ($p = .045$), 6 ($p = .058$), 7 ($p = .011$) and 9 ($p = .010$), but these differences are not significant when corrected for multiple comparisons (Bonferroni corrected $\alpha = .00167$). Additionally, when corrected for multiple comparisons, RTs for nouns are significantly faster than for verbs in all runs ($p < .001$) except for Run 3 ($p = .007$).
Figure 13. Response time on the word-level picture matching task. Response times are shown as the natural log of RT. Runs highlighted in grey were performed in the MRI scanner.
Table 4. Reaction times on the word-level picture-matching task.

<table>
<thead>
<tr>
<th>Adjective</th>
<th>Noun</th>
<th>Verb</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ms</td>
<td>log</td>
<td>ms</td>
</tr>
<tr>
<td>Run</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1155.52 (397.87)</td>
<td>6.99 (.15)</td>
<td>1157.53 (395.18)</td>
</tr>
<tr>
<td>2</td>
<td>848.71 (392.57)</td>
<td>6.64 (.29)</td>
<td>904.18 (400.03)</td>
</tr>
<tr>
<td>3</td>
<td>773.77 (389.89)</td>
<td>6.53 (.20)</td>
<td>842.74 (337.41)</td>
</tr>
<tr>
<td>4</td>
<td>759.98 (381.31)</td>
<td>6.52 (.26)</td>
<td>758.98 (313.43)</td>
</tr>
<tr>
<td>5</td>
<td>729.39 (369.49)</td>
<td>6.47 (.21)</td>
<td>769.83 (349.61)</td>
</tr>
<tr>
<td>6</td>
<td>719.66 (361.33)</td>
<td>6.47 (.25)</td>
<td>742.63 (280.69)</td>
</tr>
<tr>
<td>7</td>
<td>688.31 (367.52)</td>
<td>6.40 (.24)</td>
<td>703.21 (267.00)</td>
</tr>
<tr>
<td>8</td>
<td>648.28 (358.06)</td>
<td>6.35 (.27)</td>
<td>670.27 (268.04)</td>
</tr>
<tr>
<td>9</td>
<td>623.58 (331.40)</td>
<td>6.31 (.30)</td>
<td>668.55 (270.45)</td>
</tr>
<tr>
<td>10</td>
<td>608.42 (323.51)</td>
<td>6.31 (.34)</td>
<td>642.15 (265.57)</td>
</tr>
<tr>
<td>All</td>
<td>748.03 (393.35)</td>
<td>6.50 (.20)</td>
<td>779.97 (345.87)</td>
</tr>
</tbody>
</table>

Note: Reaction times presented in milliseconds and with a natural log transformation. Standard deviations are in parentheses.
Because significant changes in accuracy and RT were only observed during early phases of the learning process, an attempt was made to get a finer-grained picture of how early runs differed from later runs. To do this, repeated measures ANOVAs with Mini-Block (four 12-trial runs within the 48-trial runs) as a within-subjects factor were conducted on each run, collapsed across conditions. For accuracy, there was an effect of Mini-Block for Run 1, $F(3, 42) = 10.384, p < .001$, partial-$\eta^2 = .426$, observed power = .997. Bonferroni corrected post hoc comparisons revealed that accuracy was lower on Mini-Block 1 than Mini-Blocks 2 ($p = .014$), 3 ($p = .023$), and 4 ($p = .001$). There was also an effect of Mini-Block for Run 3, $F(3, 42) = 5.164, p = .004$, partial-$\eta^2 = .269$, observed power = .898, which was driven by significantly higher accuracy on Mini-Block 3 than Mini-Block 4 ($p = .042$). Accuracy across mini-blocks for the word-level picture-matching task is shown in Figure 14.
Figure 14. Accuracy on the word-level picture matching task over 40 mini-blocks, collapsed across conditions. Accuracy is shown as the proportion of correct responses. Runs highlighted in grey were performed in the MRI scanner.

For RT, there was an effect of Mini-Block for Run 1, $F(3, 42) = 22.843, p < .001$, partial-$\eta^2 = .620$, observed power = 1.00, Run 3, $F(3, 42) = 2.947, p = .044$, partial-$\eta^2 = .174$, observed power = .658, Run 5, $F(3, 42) = 3.047, p = .039$, partial-$\eta^2 = .179$, observed power = .674, and Run 6, $F(3, 42) = 6.867, p = .001$, partial-$\eta^2 = .329$, observed power = .966. For Run 1, Bonferroni corrected post hoc comparisons revealed that RT was slower on Mini-Block 1 than on Mini-Blocks 3 ($p = .006$) and 4 ($p = .001$), and that RT was slower on Mini-Block 2 than on Mini-Blocks 3 ($p = .002$) and 4 ($p = .001$). For Run 3, there was a marginally significant difference wherein RT for Mini-Block 2 was slower than 3 ($p = .058$). For Run 5, Mini-Block 1 was slower than Mini-Block 2 ($p = .025$). For Run 6, RT was slower on Mini-Block 1 than on
Mini-Blocks 2 ($p = .046$), 3 ($p = .017$), and marginally slower than Mini-Block 4 ($p = .076$). RT across mini-blocks for the word-level picture-matching task is shown in Figure 15.

![Figure 15](image)

Figure 15. Response time on the word-level picture matching task over 40 mini-blocks, collapsed across conditions. Response times are shown as the natural log of RT. Runs highlighted in grey were performed in the MRI scanner.

3.1.2 Sentence-level learning. Participants completed 8 runs of the sentence-level picture matching task over the course of three training days. Average accuracy by run and condition across all subjects is shown in Figure 16 as the proportion of correct responses, and these values, in addition to their arcsine transformations, are reported in Table 5. An $8 \times 3$ repeated measures ANOVA with Run and Condition as within-subjects factors was conducted. This revealed a main effect of Run, $F(7, 98) = 13.764, p < .001$, partial-$\eta^2 = .496$, observed
power = 1.00, a main effect of Condition, $F(2, 28) = 3.834$, $p = .034$, partial-$\eta^2 = .215$, observed power = .647, and a Run x Condition interaction, $F(14, 196) = 1.950$, $p = .024$, partial-$\eta^2 = .122$, observed power = .936. Bonferroni corrected post hoc comparisons showed that accuracy was lower on Run 1 than all other runs ($p < .005$), and that accuracy was lower on Run 2 than on Runs 7 ($p = .011$) and 8 ($p = .042$). There were no significant differences between conditions. The Run x Condition interaction was due to higher accuracy on Word Order and Lexical items than Agreement items in Runs 2 ($p = .002$ for Word Order; $p = .032$ for Lexical) and 3 ($p = .003$ for Word Order; $p = .003$ for Lexical), and no significant differences between conditions in Runs 4-8 (note that when corrected for multiple comparisons, the differences between Agreement and Lexical in Run 2 and between Agreement and Word Order in Run 3 are not significant).

Figure 16. Accuracy on the sentence-level picture matching task over the course of 8 training runs. Accuracy is shown as the proportion of correct responses.
Table 5. Accuracy on the sentence-level picture-matching task.

<table>
<thead>
<tr>
<th>Agreement</th>
<th>Lexical</th>
<th>Word Order</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion</td>
<td>Arcsine</td>
<td>Proportion</td>
<td>Arcsine</td>
</tr>
<tr>
<td>Run</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.50 (0.50)</td>
<td>0.79 (0.14)</td>
<td>0.56 (0.50)</td>
</tr>
<tr>
<td>2</td>
<td>0.53 (0.50)</td>
<td>0.82 (0.18)</td>
<td>0.68 (0.47)</td>
</tr>
<tr>
<td>3</td>
<td>0.58 (0.49)</td>
<td>0.87 (0.15)</td>
<td>0.76 (0.43)</td>
</tr>
<tr>
<td>4</td>
<td>0.70 (0.46)</td>
<td>1.02 (0.24)</td>
<td>0.75 (0.43)</td>
</tr>
<tr>
<td>5</td>
<td>0.70 (0.46)</td>
<td>1.00 (0.25)</td>
<td>0.72 (0.45)</td>
</tr>
<tr>
<td>6</td>
<td>0.74 (0.44)</td>
<td>1.10 (0.31)</td>
<td>0.79 (0.41)</td>
</tr>
<tr>
<td>7</td>
<td>0.78 (0.42)</td>
<td>1.11 (0.21)</td>
<td>0.77 (0.42)</td>
</tr>
<tr>
<td>8</td>
<td>0.76 (0.43)</td>
<td>1.12 (0.28)</td>
<td>0.76 (0.43)</td>
</tr>
<tr>
<td>All</td>
<td>0.66 (0.47)</td>
<td>0.98 (.13)</td>
<td>0.72 (0.45)</td>
</tr>
</tbody>
</table>

Note: Accuracy presented as proportion of correct responses and with an arcsine transformation. Standard deviations are in parentheses.
Average RT by run and condition is shown in Figure 17 as the natural log of RT, and raw RT data and log-transformed RT data are reported in Table 6. An 8 x 3 repeated measures ANOVA with Run and Condition as within-subjects factors was conducted. This revealed a main effect of Run, $F(7, 98) = 7.765, p < .001$, partial-$\eta^2 = .357$, observed power = 1.00, and a Run x Condition interaction, $F(14, 196) = 1.931, p = .025$, partial-$\eta^2 = .121$, observed power = .933. RT decreased over the course of time, but the only significant difference between runs after Bonferroni corrections for multiple comparisons was between Runs 1 and 8. The Run x Condition interaction appears to be due to differences in the shape of the RT curve for each condition, particularly for Runs 5 and 6 at the end of Day 2, but post hoc comparisons did not reveal any significant effects.

Figure 17. Response time on the sentence-level picture matching task. Response times are shown as the natural log of RT.
Table 6. Reaction time on the sentence-level picture-matching task.

<table>
<thead>
<tr>
<th>Agreement</th>
<th>Lexical</th>
<th>Word Order</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ms</td>
<td>log</td>
<td>ms</td>
</tr>
<tr>
<td>Run</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1874.70</td>
<td>7.50 (.25)</td>
<td>1883.42</td>
</tr>
<tr>
<td>3</td>
<td>1850.56</td>
<td>7.48 (.24)</td>
<td>1906.89</td>
</tr>
<tr>
<td>4</td>
<td>1759.91</td>
<td>7.43 (.18)</td>
<td>1813.70</td>
</tr>
<tr>
<td>5</td>
<td>1604.98</td>
<td>7.35 (.20)</td>
<td>1922.94</td>
</tr>
<tr>
<td>6</td>
<td>1690.44</td>
<td>7.37 (.22)</td>
<td>1844.43</td>
</tr>
<tr>
<td>7</td>
<td>1704.35</td>
<td>7.37 (.22)</td>
<td>1796.53</td>
</tr>
<tr>
<td>8</td>
<td>1567.94</td>
<td>7.31 (.22)</td>
<td>1649.70</td>
</tr>
<tr>
<td>All</td>
<td>1736.02</td>
<td>7.42 (.08)</td>
<td>1842.03</td>
</tr>
</tbody>
</table>

Note: Reaction times presented in milliseconds and with a natural log transformation. Standard deviations are in parentheses.
As with word learning, changes in sentence-level learning were also primarily observed during early stages of training. Therefore, performance on the sentence-level picture-matching task was also assessed by means of repeated measures ANOVAs with Mini-Block as a within-subjects factor. For accuracy, there was an effect of Mini-Block for Run 1, $F(3, 42) = 3.477, p = .024$, partial-$\eta^2 = .199$, observed power = .737, and the improvement in accuracy from Mini-Block 2 to 3 approached significance ($p = .06$). There was also an effect of Mini-Block for Run 3, $F(3, 42) = 4.077, p = .013$, partial-$\eta^2 = .226$, observed power = .809, with a significant improvement in accuracy from Mini-Block 2 to 3 ($p = .045$). Finally, there was an effect of Mini-Block for Run 8, $F(3, 42) = 3.264, p = .031$, partial-$\eta^2 = .188$, observed power = .707, due to a significant decrease in accuracy from Mini-Block 1 to 2 ($p = .020$). Accuracy across mini-blocks for the sentence-level picture-matching task is shown in Figure 18.
Figure 18. Accuracy on the sentence-level picture matching task over 32 mini-blocks, collapsed across conditions. Accuracy is shown as the proportion of correct responses.

For RT, there was a significant effect of Mini-Block for Run 2, $F(3, 42) = 3.354, p = .028$, partial-$\eta^2 = .193$, observed power = .720, driven by lower accuracy on Mini-Block 1 than 3 ($p = .030$), as well as for Run 8, $F(3, 42) = 12.173, p < .001$, partial-$\eta^2 = .465$, observed power = .999, driven by faster RT on Mini-Block 3 than on Mini-Blocks 2 ($p = .001$) and 4 ($p < .001$). RT across mini-blocks for the sentence-level picture-matching task is shown in Figure 19.
1.10 Figure 19. Response time on the sentence-level picture matching task over 32 mini-blocks, collapsed across conditions. Response times are shown as the natural log of RT.

3.1.3 Grammaticality judgment task. Accuracy and $d'$ scores on the timed GJT for Syntactic Word Order and Morphosyntactic Number Agreement sentences are presented in Table 7 and Figure 20. All statistical analyses are performed on $d'$ scores and use Day as the factor for time because the two runs within each day are combined for increased power. A 2 x 2 repeated measures ANOVA with Day and Type (Word Order, Agreement) as within-subjects factors was conducted. There was a main effect of Day, $F(1, 14) = 16.272, p = .001$, partial-$\eta^2 = .538$, observed power = .963, where $d'$ scores increased over time, a main effect of Type, $F(1, 14) = 99.993, p < .001$, partial-$\eta^2 = .877$, observed power = 1.000, with higher $d'$ scores for Word Order than Agreement, and a Day x Type interaction, $F(1, 14) = 7.682, p = .015$, partial-$\eta^2 = .354$, observed power = .732. This interaction is driven by a larger and significant increase in $d'$
scores for Word Order ($p = .001$), as opposed to Agreement, where the increase in $d'$ scores is not significant when corrected for multiple comparisons.

Table 7. Accuracy and $D'$ scores for the grammaticality judgment task.

<table>
<thead>
<tr>
<th></th>
<th>Day 1</th>
<th>Day 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>$d'$ score</td>
</tr>
<tr>
<td><strong>Morphosyntax</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammatical</td>
<td>0.67 (.47)</td>
<td>-0.54 (0.71)</td>
</tr>
<tr>
<td>Ungrammatical</td>
<td>0.18 (.39)</td>
<td></td>
</tr>
<tr>
<td><strong>Syntax</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammatical</td>
<td>0.72 (.45)</td>
<td>0.91 (0.91)</td>
</tr>
<tr>
<td>Ungrammatical</td>
<td>0.57 (.49)</td>
<td>0.83 (.38)</td>
</tr>
</tbody>
</table>

Note: Average accuracy, reported as the proportion of correct responses and $d'$ scores, on the violation task on Day 1 and Day 3. Standard deviations are in parentheses.
Figure 20. D’ values for the grammaticality judgment task.

Reaction times in milliseconds and with natural log transformations for the GJT are shown in Table 8. RTs are reported separately for hits, false alarms, correct rejections, and misses, and these were analyzed in four different ways, as shown in Figure 21. Four separate 2 x 2 x 2 repeated measures ANOVAs were run, with Day, Type, and Grammaticality (Grammatical, Ungrammatical) as within-subjects factors. The first ANOVA examined correct responses only, which is a typical analysis for RT data. These included hits, where the subject chose “good” for a grammatical sentence, and correct rejections, where a subject chose “bad” for an ungrammatical sentence. There was a marginally significant main effect of Day, $F(1,11) = 3.818, p = .077$, partial-$\eta^2 = .258$, observed power = .430, wherein RT was faster on Day 3, and a main effect of Type, $F(1,11) = 11.978, p = .005$, partial-$\eta^2 = .521$, observed power = .882, where RT for Word Order was faster than for Agreement. Additionally, there was a Day x Type interaction, $F(1,11) = 13.696, p = .003$, partial-$\eta^2 = .555$, observed power = .920, a Day x Grammaticality
interaction: $F(1, 11) = 15.700, p = .002$, partial-$\eta^2 = .588$, observed power = .949, and a Type x Grammaticality interaction: $F(1, 11) = 23.359, p = .001$, partial-$\eta^2 = .680$, observed power = .992.

Regarding the Day x Type interaction, RT was faster on Day 3 than Day 1 for Word Order sentences ($p = .004$ for correct rejections, and marginally significant at $p = .053$ for hits), but there was no change for Agreement sentences. Regarding the Day x Grammaticality interaction, the difference in RT between Grammatical (hits) and Ungrammatical (correct rejections) sentences was large and significant in Day 3 ($p < .001$) but not Day 1 when corrected for multiple comparisons. Unlike for word order, there were no differences between Agreement sentences according to Grammaticality, which appears to have driven the Day x Grammaticality interaction.

The second ANOVA examined incorrect responses only. These included misses, where a subject chose “bad” for a grammatical sentence, and false alarms, where a subject chose “good” for an ungrammatical sentence. There were no significant effects from this ANOVA.

The next set of ANOVAs examined endorsements and rejections. Up until this point, $d'$ and RT analyses on the GJT have shown that subjects were unable to distinguish grammatical and ungrammatical agreement sentences. However, accuracy is a rather coarse measure for determining psychological processes, and the RT analyses on correct and incorrect sentences confound grammaticality and type of response. That is, in the “correct” analysis (Figure 21a), trials where the sentence was grammatical and the subject chose “good” were compared to trials where the sentence was ungrammatical and the subject chose “bad.” It is well-established in the psychology literature that response latencies vary as a function of type of response, like “yes” and “no” (Pike, 1973). By looking at only endorsements and only rejections, we can test the null
hypothesis that subjects are processing all agreement sentences in the same way. If, for example, RT is slower when subjects are endorsing ungrammatical sentences than when they are endorsing grammatical sentences, this would suggest that they might be hesitating to (incorrectly) endorse ungrammatical sentences based on something that they have learned in training.

The third ANOVA examined endorsements only. These included any trials where the subject responded “good” – hits and false alarms. This revealed a significant main effect of Grammaticality, \( F(1,12) = 4.777, p = .049 \), partial-\( \eta^2 = .285 \), observed power = .520, where grammatical sentences were responded to more quickly than ungrammatical sentences (\( p = .049 \)). While there were no significant differences in Day 1, RT on Day 3 was significantly faster for endorsements of matched Grammatical sentences than for Morphosyntactic Violations (\( p = .003 \)).

The fourth ANOVA examined rejections only. These included any trials where the subject responded “bad” – misses and correct rejections. This revealed a marginally significant main effect of Day, \( F(1,11) = 3.495, p = .088 \), partial-\( \eta^2 = .241 \), observed power = .400, where RT decreased in Day 3, a main effect of Type, \( F(1,11) = 11.233, p = .006 \), partial-\( \eta^2 = .505 \), observed power = .862, where RT was faster for Word Order than for Agreement, and a main effect of Grammaticality, \( F(1,11) = 14.532, p = .003 \), partial-\( \eta^2 = .569 \), observed power = .934, where RTs for ungrammatical sentences were faster than for grammatical sentences. In addition, there was a Day x Grammaticality interaction, \( F(1,11) = 9.678, p = .010 \), partial-\( \eta^2 = .468 \), observed power = .808 and a Type x Grammaticality interaction, \( F(1,11) = 15.983, p = .002 \), partial-\( \eta^2 = .592 \), observed power = .953. The three way interaction of Day x Type x
Grammaticality was marginally significant, $F(1,11) = 3.574$, $p = .085$, partial-$\eta^2 = .245$, observed power $= .407$. The Day x Grammaticality interaction was likely driven by faster RTs for correct rejections than for misses for Word Order sentences in Day 3, but not Day 1. The Type x Grammaticality interaction appears to be due to differences between grammatical and ungrammatical items for Word Order, but not for Agreement. However, no post hoc tests for these interactions were significant.
Table 8. Reaction time for grammaticality judgment task.

<table>
<thead>
<tr>
<th></th>
<th>Day 1</th>
<th>Day 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Morphosyntax</td>
<td>Syntax</td>
</tr>
<tr>
<td><strong>Hit</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ms</td>
<td>861.46 (525.10)</td>
<td>929.71 (537.20)</td>
</tr>
<tr>
<td>log</td>
<td>6.57 (0.61)</td>
<td>6.66 (0.61)</td>
</tr>
<tr>
<td><strong>False Alarm</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ms</td>
<td>848.23 (514.24)</td>
<td>896.95 (519.70)</td>
</tr>
<tr>
<td>log</td>
<td>6.56 (0.61)</td>
<td>6.62 (0.62)</td>
</tr>
<tr>
<td><strong>Correct Rejection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ms</td>
<td>1030.51 (545.05)</td>
<td>813.14 (495.15)</td>
</tr>
<tr>
<td>log</td>
<td>6.77 (0.62)</td>
<td>6.52 (0.62)</td>
</tr>
<tr>
<td><strong>Miss</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ms</td>
<td>1108.03 (612.12)</td>
<td>1026.03 (555.78)</td>
</tr>
<tr>
<td>log</td>
<td>6.84 (0.61)</td>
<td>6.79 (0.55)</td>
</tr>
</tbody>
</table>

Note: Reaction times presented in milliseconds and with a natural log transformation. Standard deviations are in parentheses.
Figure 21. Response time on the grammaticality judgment task on Days 1 and 3. Response times are shown as the natural log of RT.

A) Correct responses only; B) incorrect responses only; C) endorsements only; D) rejections only. FA = False alarm; CR = Correct rejection.
3.1.4 Production. The average numbers of correctly pronounced words (out of 24) are presented in Table 9 and Figure 22. A repeated measures ANOVA revealed a significant effect of Day, $F(2, 24) = 33.780, p < .001$. Bonferroni corrected post hoc comparisons showed that performance increased significantly between each day (all $ps < .005$).

Table 9. Performance on the word production task.

<table>
<thead>
<tr>
<th>Day</th>
<th>Number Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.31 (4.64)</td>
</tr>
<tr>
<td>2</td>
<td>11.69 (5.58)</td>
</tr>
<tr>
<td>3</td>
<td>15.69 (6.21)</td>
</tr>
</tbody>
</table>

Note: Mean performance across subjects. SD in parentheses.

Figure 22. Performance on the word production task.
The average overall score on the sentence production task (out of 15.5) for each day are shown in Table 10 and Figure 23. A repeated measures ANOVA revealed a significant effect of Day, $F(2, 28) = 51.974, p < .001$, partial $\eta^2 = .788$, observed power = 1.0. Bonferroni corrected post hoc comparisons showed that performance increased significantly between each day (all $p$s < .001).

Table 10. Performance on the sentence production task.

<table>
<thead>
<tr>
<th>Day</th>
<th>Number Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.75 (2.94)</td>
</tr>
<tr>
<td>2</td>
<td>10.92 (2.32)</td>
</tr>
<tr>
<td>3</td>
<td>12.51 (2.08)</td>
</tr>
</tbody>
</table>

Note: Mean performance across subjects. SD in parentheses.

Figure 23. Performance on the sentence production task.
3.2 fMRI Results

3.2.1 Whole Brain.

3.2.1.1 Word-level. At the word-level, only a single voxel in one contrast was significant. On Day 3, words activated a voxel adjacent to the right posterior hippocampus (42, -38, 0) when contrasted with baseline.

3.2.1.2 Sentence-level. At the sentence-level, six contrasts resulted in significant activation. These included the contrast of the sentence task versus baseline (and vice versa) collapsed across all eight runs, as well as these same contrasts on Day 1 (i.e., Runs 1 and 2) and Day 3 (i.e., Runs 7 and 8). Peak contrasts that survived a FWE correction at $p < .05$ are displayed in Tables 11, 12, 13, and Figure 24.
Table 11. Peaks for contrasts between the sentence learning task and baseline.

<table>
<thead>
<tr>
<th>Anatomical Location</th>
<th>BA</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>k_e</th>
<th>T</th>
<th>Z_e</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentence &gt; Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Inferior Frontal Gyrus</td>
<td>45</td>
<td>-48</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>12.78</td>
<td>5.58</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>-44</td>
<td>8</td>
<td>32</td>
<td>1</td>
<td>10.61</td>
<td>5.21</td>
</tr>
<tr>
<td>L Inferior Parietal Lobule</td>
<td>7</td>
<td>-32</td>
<td>-60</td>
<td>44</td>
<td>1</td>
<td>9.95</td>
<td>5.08</td>
</tr>
<tr>
<td>L Middle Occipital Lobule</td>
<td>19</td>
<td>-30</td>
<td>-72</td>
<td>34</td>
<td>11</td>
<td>13.18</td>
<td>5.64</td>
</tr>
<tr>
<td>R Inferior Frontal Gyrus</td>
<td>45</td>
<td>54</td>
<td>32</td>
<td>28</td>
<td>1</td>
<td>10.56</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>42</td>
<td>22</td>
<td>24</td>
<td>91</td>
<td>19.80</td>
<td>6.40</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>4</td>
<td>14</td>
<td>54</td>
<td>19</td>
<td>12.19</td>
<td>5.49</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>48</td>
<td>8</td>
<td>38</td>
<td>5</td>
<td>11.40</td>
<td>5.36</td>
</tr>
<tr>
<td>R Inferior Parietal Lobe</td>
<td>7</td>
<td>34</td>
<td>-54</td>
<td>50</td>
<td>5</td>
<td>10.46</td>
<td>5.18</td>
</tr>
<tr>
<td>R Angular Gyrus</td>
<td>39</td>
<td>30</td>
<td>-62</td>
<td>44</td>
<td>22</td>
<td>14.36</td>
<td>5.81</td>
</tr>
<tr>
<td><strong>Baseline &gt; Sentence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Prefrontal Cortex</td>
<td>10</td>
<td>0</td>
<td>66</td>
<td>6</td>
<td>2</td>
<td>10.13</td>
<td>5.12</td>
</tr>
<tr>
<td>L Cingulate Gyrus</td>
<td>24</td>
<td>-6</td>
<td>48</td>
<td>-16</td>
<td>3</td>
<td>9.93</td>
<td>5.08</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>-4</td>
<td>58</td>
<td>22</td>
<td>1</td>
<td>9.92</td>
<td>5.07</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>-8</td>
<td>40</td>
<td>-12</td>
<td>32</td>
<td>12.85</td>
<td>5.59</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>6</td>
<td>44</td>
<td>-14</td>
<td>2</td>
<td>9.96</td>
<td>5.08</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-12</td>
<td>-8</td>
<td>44</td>
<td>2</td>
<td>10.81</td>
<td>5.25</td>
</tr>
<tr>
<td>Anatomical Location</td>
<td>BA</td>
<td>x</td>
<td>y</td>
<td>z</td>
<td>ke</td>
<td>T</td>
<td>Z_e</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>-6</td>
<td>-16</td>
<td>36</td>
<td></td>
<td>10.58</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-4</td>
<td>-24</td>
<td>40</td>
<td>18</td>
<td>11.00</td>
<td>5.28</td>
</tr>
<tr>
<td>L inferior parietal lobule</td>
<td>7</td>
<td>-54</td>
<td>-42</td>
<td>32</td>
<td>1</td>
<td>10.26</td>
<td>5.14</td>
</tr>
<tr>
<td>L angular gyrus</td>
<td>39</td>
<td>-58</td>
<td>-56</td>
<td>26</td>
<td>18</td>
<td>11.95</td>
<td>5.45</td>
</tr>
<tr>
<td>R medial frontal gyrus</td>
<td>9</td>
<td>8</td>
<td>54</td>
<td>30</td>
<td>4</td>
<td>10.19</td>
<td>5.13</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>24</td>
<td>46</td>
<td>44</td>
<td>13</td>
<td>13.42</td>
<td>5.68</td>
</tr>
<tr>
<td>R anterior cingulate cortex</td>
<td>24</td>
<td>2</td>
<td>32</td>
<td>-12</td>
<td>14</td>
<td>11.00</td>
<td>5.28</td>
</tr>
<tr>
<td>R temporal pole</td>
<td>38</td>
<td>44</td>
<td>8</td>
<td>-32</td>
<td>3</td>
<td>12.25</td>
<td>5.50</td>
</tr>
<tr>
<td>R postcentral gyrus</td>
<td>3</td>
<td>36</td>
<td>-16</td>
<td>28</td>
<td>2</td>
<td>10.93</td>
<td>5.27</td>
</tr>
<tr>
<td>R rolandic operculum</td>
<td>1</td>
<td>46</td>
<td>-22</td>
<td>22</td>
<td>1</td>
<td>9.88</td>
<td>5.07</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>26</td>
<td>-26</td>
<td>18</td>
<td>1</td>
<td>10.18</td>
<td>5.13</td>
</tr>
<tr>
<td>R cingulate gyrus</td>
<td>6</td>
<td>6</td>
<td>-26</td>
<td>46</td>
<td>3</td>
<td>10.26</td>
<td>5.14</td>
</tr>
</tbody>
</table>

Note: Clusters are significant at $p < .05$ (FWE).
Table 12. Peaks for contrasts between the sentence learning task and baseline on Day 1.

<table>
<thead>
<tr>
<th>Anatomical Location</th>
<th>BA</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>k_e</th>
<th>T</th>
<th>Z_e</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentence &gt; Baseline (Day 1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R inferior frontal gyrus</td>
<td>45</td>
<td>46</td>
<td>24</td>
<td>24</td>
<td>42</td>
<td>12.34</td>
<td>5.66</td>
</tr>
<tr>
<td><strong>Baseline &gt; Sentence (Day 1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R prefrontal cortex</td>
<td>9</td>
<td>-20</td>
<td>58</td>
<td>30</td>
<td>1</td>
<td>9.37</td>
<td>5.08</td>
</tr>
</tbody>
</table>

Note: Clusters are significant at $p < .05$ (FWE).
Figure 24. Sentence-level activation in the whole-brain analysis. $p < .05$, FWE correction.
Table 13. Peaks for contrasts between the sentence learning task and baseline on Day 3.

<table>
<thead>
<tr>
<th>Anatomical Location</th>
<th>BA</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>$k_e$</th>
<th>$T$</th>
<th>$Z_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentence &gt; Baseline (Day 3)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L middle occipital lobule</td>
<td>19</td>
<td>-28</td>
<td>-74</td>
<td>32</td>
<td>11</td>
<td>10.56</td>
<td>5.33</td>
</tr>
<tr>
<td>R inferior frontal gyrus</td>
<td>45</td>
<td>46</td>
<td>22</td>
<td>28</td>
<td>57</td>
<td>12.64</td>
<td>5.71</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>46</td>
<td>10</td>
<td>34</td>
<td>1</td>
<td>9.20</td>
<td>5.04</td>
</tr>
<tr>
<td>R precentral gyrus</td>
<td>6</td>
<td>4</td>
<td>12</td>
<td>52</td>
<td>9</td>
<td>9.59</td>
<td>5.13</td>
</tr>
<tr>
<td><strong>Baseline &gt; Sentence (Day 3)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L anterior cingulate cortex</td>
<td>24</td>
<td>-4</td>
<td>34</td>
<td>2</td>
<td>5</td>
<td>10.36</td>
<td>5.29</td>
</tr>
<tr>
<td>L supramarginal gyrus</td>
<td>40</td>
<td>-54</td>
<td>-40</td>
<td>32</td>
<td>5</td>
<td>10.51</td>
<td>5.33</td>
</tr>
<tr>
<td>L insula</td>
<td></td>
<td>-40</td>
<td>-14</td>
<td>-8</td>
<td>2</td>
<td>9.71</td>
<td>5.16</td>
</tr>
<tr>
<td>R prefrontal cortex</td>
<td>9</td>
<td>10</td>
<td>56</td>
<td>10</td>
<td>2</td>
<td>9.60</td>
<td>5.13</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>10</td>
<td>52</td>
<td>28</td>
<td>1</td>
<td>9.21</td>
<td>5.04</td>
</tr>
<tr>
<td>R anterior cingulate cortex</td>
<td>24</td>
<td>2</td>
<td>34</td>
<td>-6</td>
<td>2</td>
<td>9.36</td>
<td>5.08</td>
</tr>
</tbody>
</table>

Note: Clusters are significant at $p < .05$ (FWE).

**3.2.1.3 Sentence conditions.** When each sentence condition was analyzed separately over all runs, there was significant activation for all contrasts in both directions. Peak contrasts that survived a FWE correction at $p < .05$ are displayed in Tables 14, 15, and 16.
Table 14. Peaks for contrasts between the lexical condition and baseline over all runs.

<table>
<thead>
<tr>
<th>Anatomical Location</th>
<th>BA</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>k_e</th>
<th>T</th>
<th>Z_e</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lexical &gt; Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L middle frontal gyrus</td>
<td>10</td>
<td>-36</td>
<td>64</td>
<td>4</td>
<td>3</td>
<td>10.32</td>
<td>5.15</td>
</tr>
<tr>
<td>L inferior parietal lobule</td>
<td>7</td>
<td>-48</td>
<td>-50</td>
<td>38</td>
<td>47</td>
<td>17.20</td>
<td>6.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-52</td>
<td>-56</td>
<td>32</td>
<td></td>
<td>11.01</td>
<td>5.29</td>
</tr>
<tr>
<td>L inferior occipital lobule</td>
<td>19</td>
<td>-32</td>
<td>-80</td>
<td>-8</td>
<td>4</td>
<td>10.42</td>
<td>5.17</td>
</tr>
<tr>
<td>R supramarginal gyrus</td>
<td>40</td>
<td>54</td>
<td>-36</td>
<td>34</td>
<td>1</td>
<td>10.11</td>
<td>5.11</td>
</tr>
<tr>
<td>R cerebellum</td>
<td>14</td>
<td>-78</td>
<td>-32</td>
<td>1</td>
<td></td>
<td>9.97</td>
<td>5.08</td>
</tr>
<tr>
<td>R putamen</td>
<td>28</td>
<td>0</td>
<td>6</td>
<td>16</td>
<td></td>
<td>13.26</td>
<td>5.65</td>
</tr>
<tr>
<td><strong>Baseline &gt; Lexical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L superior temporal gyrus</td>
<td>22</td>
<td>-60</td>
<td>-10</td>
<td>2</td>
<td>7</td>
<td>10.85</td>
<td>5.26</td>
</tr>
<tr>
<td>R superior temporal gyrus</td>
<td>22</td>
<td>64</td>
<td>-6</td>
<td>8</td>
<td>4</td>
<td>11.03</td>
<td>5.29</td>
</tr>
</tbody>
</table>

Note: Clusters are significant at \( p < .05 \) (FWE).
Table 15. Peaks for contrasts between the agreement condition and baseline over all runs.

<table>
<thead>
<tr>
<th>Anatomical Location</th>
<th>BA</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>$k_e$</th>
<th>$T$</th>
<th>$Z_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Agreement &gt; Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L inferior parietal lobule</td>
<td>7</td>
<td>-40</td>
<td>-58</td>
<td>50</td>
<td>1</td>
<td>11.14</td>
<td>5.31</td>
</tr>
<tr>
<td>R inferior frontal gyrus</td>
<td>47</td>
<td>46</td>
<td>52</td>
<td>-12</td>
<td>1</td>
<td>10.88</td>
<td>5.26</td>
</tr>
<tr>
<td>R putamen</td>
<td>28</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>10.83</td>
<td>5.25</td>
</tr>
<tr>
<td><strong>Baseline &gt; Agreement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L rolandic operculum</td>
<td>4</td>
<td>-64</td>
<td>-6</td>
<td>12</td>
<td>2</td>
<td>11.09</td>
<td>5.30</td>
</tr>
<tr>
<td>L superior temporal gyrus</td>
<td>22</td>
<td>-58</td>
<td>-12</td>
<td>4</td>
<td>11</td>
<td>12.83</td>
<td>5.59</td>
</tr>
<tr>
<td>L paracentral lobule</td>
<td>4</td>
<td>-4</td>
<td>-26</td>
<td>56</td>
<td>5</td>
<td>11.26</td>
<td>5.33</td>
</tr>
<tr>
<td>R superior temporal gyrus</td>
<td>22</td>
<td>54</td>
<td>-12</td>
<td>8</td>
<td>4</td>
<td>10.66</td>
<td>5.22</td>
</tr>
</tbody>
</table>

Note: Clusters are significant at $p < .05$ (FWE).
Table 16. Peaks for contrasts between the word-order condition and baseline over all runs

<table>
<thead>
<tr>
<th>Anatomical Location</th>
<th>BA</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>$k_e$</th>
<th>$T$</th>
<th>$Z_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word order &gt; Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L superior frontal gyrus</td>
<td>10</td>
<td>64</td>
<td>2</td>
<td>7</td>
<td>11.07</td>
<td>5.30</td>
<td></td>
</tr>
<tr>
<td>L middle frontal gyrus</td>
<td>9</td>
<td>-46</td>
<td>36</td>
<td>30</td>
<td>1</td>
<td>9.95</td>
<td>5.08</td>
</tr>
<tr>
<td>L inferior parietal lobule</td>
<td>7</td>
<td>-46</td>
<td>-50</td>
<td>38</td>
<td>15</td>
<td>13.46</td>
<td>5.68</td>
</tr>
<tr>
<td>L angular gyrus</td>
<td>39</td>
<td>-50</td>
<td>-56</td>
<td>32</td>
<td>14</td>
<td>12.33</td>
<td>5.51</td>
</tr>
<tr>
<td>L precuneus</td>
<td>7</td>
<td>-12</td>
<td>-66</td>
<td>32</td>
<td>1</td>
<td>9.88</td>
<td>5.07</td>
</tr>
<tr>
<td>L middle occipital lobule</td>
<td>19</td>
<td>-30</td>
<td>-72</td>
<td>36</td>
<td>8</td>
<td>10.68</td>
<td>5.22</td>
</tr>
<tr>
<td>L inferior occipital lobule</td>
<td>10</td>
<td>-32</td>
<td>-78</td>
<td>-6</td>
<td>11</td>
<td>13.58</td>
<td>5.70</td>
</tr>
<tr>
<td>L insula</td>
<td></td>
<td>-32</td>
<td>18</td>
<td>6</td>
<td>5</td>
<td>11.42</td>
<td>5.36</td>
</tr>
<tr>
<td>L thalamus</td>
<td></td>
<td>-16</td>
<td>-30</td>
<td>4</td>
<td>2</td>
<td>11.44</td>
<td>5.36</td>
</tr>
<tr>
<td>R supramarginal gyrus</td>
<td>40</td>
<td>54</td>
<td>-38</td>
<td>34</td>
<td>2</td>
<td>12.10</td>
<td>5.47</td>
</tr>
<tr>
<td>R inferior parietal lobule</td>
<td>7</td>
<td>48</td>
<td>-50</td>
<td>46</td>
<td>25</td>
<td>11.64</td>
<td>5.40</td>
</tr>
<tr>
<td>R insula</td>
<td></td>
<td>34</td>
<td>18</td>
<td>8</td>
<td>1</td>
<td>9.89</td>
<td>5.07</td>
</tr>
<tr>
<td>R putamen</td>
<td></td>
<td>28</td>
<td>2</td>
<td>4</td>
<td>37</td>
<td>16.13</td>
<td>6.02</td>
</tr>
<tr>
<td>R cerebellum</td>
<td></td>
<td>26</td>
<td>-60</td>
<td>-32</td>
<td>4</td>
<td>10.96</td>
<td>5.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>-78</td>
<td>-32</td>
<td>19</td>
<td>12.17</td>
<td>5.49</td>
</tr>
<tr>
<td><strong>Baseline &gt; Word order</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L anterior cingulate cortex</td>
<td>24</td>
<td>0</td>
<td>38</td>
<td>-6</td>
<td>6</td>
<td>10.25</td>
<td>5.14</td>
</tr>
<tr>
<td>L superior temporal gyrus</td>
<td>22</td>
<td>-58</td>
<td>-12</td>
<td>4</td>
<td>7</td>
<td>12.01</td>
<td>5.46</td>
</tr>
</tbody>
</table>
3.2.1.4 Timed GJT. There were no suprathreshold clusters in any of the contrasts for the timed GJT at the whole-brain level.

3.2.2 Region of interest (ROI) analyses. There were few contrasts that resulted in significant activation at the whole-brain level, especially for the word-level picture-matching task and the GJT. ROI analyses allow for the investigation of activation in specific, theoretically motivated regions, with greater power.

3.2.2.1 Word-level. When both word-level picture-matching runs were collapsed, ROI analyses revealed significant activation in right BA45, $t(13) = 2.233, p = .044$, and a trend toward significant activation in right BA44, $t(13) = 1.988, p = .068$. There was significant deactivation in the right putamen, $t(13) = -2.508, p = .026$. The same pattern of activation and deactivation was apparent when Day 3 was analyzed separately, but no regions were significant (R BA45: $p = .096$; R BA44: $p = .071$; R putamen: $p = .097$). On Day 1, the right hippocampus was activated, but not significantly, $t(13) = 1.875, p = .083$. Left BA44, $t(13) = -3.103, p = .008$, and left BA6, $t(13) = -2.558, p = .024$, were significantly deactivated, and left BA45, $t(13) = -1.852, p = .087$, and right BA6, $t(13) = -1.887, p = .082$, were also deactivated, but not significantly. There was a significant increase in percent signal change from Day 1 to Day 3 in left BA44, $t(13) = 3.432, p = .004$, and left BA6, $t(13) = 2.433, p = .030$, but no significant decreases (see Figure 26). Trending activation on Day 1 and Day 3 are shown in Figure 25.
Figure 25. Word-level activation in ROI analyses, showing right hippocampal activation on Day 1 and LIFG on Day 3. The images show results from a small volume correction of all 16 ROIs combined at $p < .05$.

Figure 26. Changes in activation in (a) L BA44 and (b) L BA6 at the word level.

3.2.2.2 Sentence-level. For the sentence-baseline contrast, right BA44 and bilateral BA45 were activated. Right caudate, bilateral BA47, bilateral hippocampus and parahippocampal gyrus, and bilateral putamen were deactivated. In the feedback-baseline contrast, bilateral BA44, left BA47, bilateral BA45, and bilateral putamen were activated. Bilateral hippocampus and
parahippocampal gyrus were deactivated. Significance levels and $t$-statistics for all regions are listed in Table 17.

One-way ANOVAs revealed that there were no main effects of time for any of the ROIs. That is, none of these ROIs exhibited significant changes in activation over the course of learning.
Table 17. ROI analyses for the sentence-level picture-matching task.

<table>
<thead>
<tr>
<th>Anatomical Location</th>
<th>Sentence</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
<td>$p$</td>
</tr>
<tr>
<td>$L$ hippocampus</td>
<td>-5.256</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>$R$ hippocampus</td>
<td>-4.664</td>
<td>.001</td>
</tr>
<tr>
<td>$L$ parahippocampal gyrus</td>
<td>-4.279</td>
<td>.001</td>
</tr>
<tr>
<td>$R$ parahippocampal gyrus</td>
<td>-4.189</td>
<td>.001</td>
</tr>
<tr>
<td>$L$ caudate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R$ caudate</td>
<td>-2.228</td>
<td>.046</td>
</tr>
<tr>
<td>$L$ putamen</td>
<td>-4.982</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>$R$ putamen</td>
<td>-4.209</td>
<td>.001</td>
</tr>
<tr>
<td>$L$ BA6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R$ BA6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L$ BA44</td>
<td>4.941</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>$R$ BA44</td>
<td>5.513</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>$L$ BA45</td>
<td>2.353</td>
<td>.037</td>
</tr>
<tr>
<td>$R$ BA45</td>
<td>5.939</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>$L$ BA47</td>
<td>-3.641</td>
<td>.003</td>
</tr>
<tr>
<td>$R$ BA47</td>
<td>-3.055</td>
<td>.010</td>
</tr>
</tbody>
</table>

Note: Activation is significant at $p < .05$.  

132
3.2.2.3 **Timed GJT.** When all runs of the timed GJT were collapsed for the contrast of syntactic violations minus correct sentences, ROI analyses revealed significant activation in the left IFG, including BA44, \( t(13) = 2.660, p = .020 \) and BA45, \( t(13) = 2.227, p = .044 \), and a trend toward significant activation in BA47, \( t(13) = 1.964, p = .071 \). Right BA47 was also significantly activated, \( t(13) = 3.033, p = .010 \). There was a trend toward significant deactivation in BA6, \( t(13) = -2.018, p = .065 \). The pattern for bilateral activation in BA47 (L: \( t(13) = 1.806, p = .094 \); R: \( t(13) = 2.455, p = .029 \)), and deactivation in right BA6 was also found in Day 1, though the deactivation reached significance, \( t(13) = -2.744, p = .017 \). Right BA47 was also activated in Day 3, \( t(13) = 2.865, p = .013 \), as was left BA44, \( t(13) = 2.740, p = .017 \). There was a trend toward significant activation in bilateral BA45 (L: \( t(13) = 2.072, p = .059 \); R: \( t(13) = 2.048, p = .061 \)). No areas were deactivated on Day 3. There was a marginally significant increase in percent signal change from Day 1 to Day 3 in the right putamen, \( t(13) = 1.977, p = .070 \), but no decreases (see Figure 28). There were no significant activations or deactivations for any ROIs for the contrast of morphosyntactic violations versus correct sentences. 

A small volume correction analysis was also run on the 16 ROIs combined for the contrast of syntactic violations versus grammatical sentences in Day 3 versus Day 1. This revealed significant activation \( (p < .01) \) in the left caudate \( (x = -16, y = -8, z = 22, Z_e = 2.99) \) and the right putamen \( (x = 32, y = -6, z = -4, Z_e = 3.00) \). See Figure 27 for activation related to the GJT.
Figure 27. Activation for syntactic violations versus grammatical sentences from ROI and SVC analyses. The images are based on a SVC of the 16 combined ROIs at $p < .01$.

Figure 28. Changes in activation in the right putamen for syntactic violations.

3.2.3 The relationship between the brain and behavior. The above analyses provide information on the relationship between the amount of exposure subjects have to a language and
changes in neural activity. However, they cannot directly tell us about the relationship between learning and neural activity. Analyses that use exposure as a metric for gauging learning-related neural changes assume that learning is uniform across runs and across subjects. Behavioral results (see Section 3.1) clearly demonstrate that learning happens very quickly when subjects are first exposed to the language, but improvements in behavioral performance taper off as exposure increases. Furthermore, there is an enormous range of individual differences among the 15 subjects in this study, particularly in the sentence-level picture-matching task and the GJT, though this is even evident in the first word-level picture-matching run. See Appendix F for figures depicting individual performance in the three fMRI tasks. The assumption that exposure is linearly related to increases in proficiency across time and across subjects is not met in this study. In order to explore the relationship between learning and changes in neural activity, it is necessary to use a more direct measure of learning. In this case, that is behavioral performance.

3.2.3.1 Word-level picture-matching. At the whole-brain level, there were no significant areas of activation ($p < .05$, FWE) on Day 1 or Day 3 predicted by the arcsine of word accuracy in Run 1. At the ROI level, there were no significant correlations between accuracy and activation for either day. However, there were several marginally significant positive correlations with moderate effect sizes between the arcsine of accuracy for Run 1 and activation in the left parahippocampal gyrus ($r = .552, p = .040$), left hippocampus ($r = .490, p = .075$), left BA44 ($r = .480, p = .082$), and bilateral BA47 (left: $r = .462, p = .096$; right: $r = .505, p = .066$), on Day 1 (Bonferroni corrected $p < .003$). Scatterplots of these correlations are shown in Figure 29.
3.2.3.2 Sentence-level picture-matching. At the whole-brain level, there were no significant areas of activation ($p < .05$, FWE) on Day 1 or Day 3 predicted by the arcsine of accuracy in Run 1, Run 8, Run$_{MAX}$, Delta$_{8-1}$, or Delta$_{MAX}$. At the ROI level, the arcsine of accuracy for each run within subjects was not correlated with activation or deactivation in any of the ROIs. Pearson’s correlations between the arcsine of accuracy and activation across subjects for each run revealed a negative relationship between performance and activation in the hippocampus and parahippocampal gyrus in Run 1, and increased activation in frontal regions,

Figure 29. Scatterplots of the beta values against the arcsine of accuracy for the first word-level picture-matching run.
particularly left BA6, in Run 5. Though not significant when corrected for multiple comparisons, there is a pattern of a sustained negative correlation between medial temporal regions and activation in Run 3, and evidence of a shift in IFG structures across runs. All significant and trending \((p < .1)\) correlations are shown in Table 18.

Table 18. Significant and trending correlations between accuracy and activation across runs.

<table>
<thead>
<tr>
<th>Region</th>
<th>Run 1</th>
<th>Run 3</th>
<th>Run 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(r)</td>
<td>(p)</td>
<td>(r)</td>
</tr>
<tr>
<td>(L) hippocampus</td>
<td>-.747</td>
<td>.003*</td>
<td></td>
</tr>
<tr>
<td>(R) hippocampus</td>
<td>-.708</td>
<td>.007</td>
<td>-.505</td>
</tr>
<tr>
<td>(L) parahippocampal gyrus</td>
<td>-.789</td>
<td>.001*</td>
<td>-.531</td>
</tr>
<tr>
<td>(R) parahippocampal gyrus</td>
<td>-.778</td>
<td>.002*</td>
<td>-.550</td>
</tr>
<tr>
<td>(L) BA47</td>
<td>-.549</td>
<td>.052</td>
<td></td>
</tr>
<tr>
<td>(R) BA47</td>
<td>-.547</td>
<td>.053</td>
<td></td>
</tr>
<tr>
<td>(L) BA45</td>
<td></td>
<td></td>
<td>-.532</td>
</tr>
<tr>
<td>(L) BA6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R) BA6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(L) BA44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * indicates significant at Bonferroni corrected \(p < .003\).

\subsection*{3.2.3.3 Timed GJT} At the whole-brain level, there were no significant areas of activation \((p < .05, \text{FWE})\) on Day 1 or Day 3 that correlated with \(d’\) scores on Day 1, Day 3, or the Delta_{3-1}
for either morphosyntactic or syntactic violations. At the ROI level, there were no significant correlations between Delta$_{3-1}$ and activation for Day 1 or Day 3 when a Bonferroni correction for multiple corrections was applied ($p < .003$). However, there were some marginally significant trends with moderate effect sizes. For morphosyntax, there was a positive correlation between Delta$_{3-1}$ and activation in the right hippocampus on Day 1 and a negative correlation with the left parahippocampal gyrus on Day 3. For syntax, several regions in bilateral IFG were positively correlated with performance on Day 3, as well as right caudate and left putamen. These and any other marginally significant ($p < .1$) correlations are shown in Table 19.
Table 19. Correlations between delta $d'$ values and activation for syntactic and morphosyntactic violations.

<table>
<thead>
<tr>
<th>Region</th>
<th>Correlations for syntactic trials</th>
<th>Correlations for morphosyntactic trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1</td>
<td>Day 3</td>
</tr>
<tr>
<td></td>
<td>$r$</td>
<td>$p$</td>
</tr>
<tr>
<td>$R$ hippocampus</td>
<td>.548</td>
<td>.043</td>
</tr>
<tr>
<td>$L$ parahippocampal gyrus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R$ caudate</td>
<td>.460</td>
<td>.098</td>
</tr>
<tr>
<td>$L$ putamen</td>
<td>.722</td>
<td>.004</td>
</tr>
<tr>
<td>$L$ BA47</td>
<td>.476</td>
<td>.085</td>
</tr>
<tr>
<td>$R$ BA47</td>
<td>.573</td>
<td>.032</td>
</tr>
<tr>
<td>$R$ BA45</td>
<td>.632</td>
<td>.015</td>
</tr>
<tr>
<td>$L$ BA6</td>
<td>.586</td>
<td>.028</td>
</tr>
<tr>
<td>$R$ BA6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R$ BA44</td>
<td>.524</td>
<td>.054</td>
</tr>
</tbody>
</table>
Chapter 4: Discussion & Conclusions

4.1 Behavioral Results

Consistent with previous research suggesting that word-learning happens rapidly with few exemplars, the Mini-Basque learners reached ceiling for accuracy on the word-level picture-matching task after the first run of training (i.e., four exemplars per word), and for RT after the second run. By the time learners started sentence-level training in the MRI scanner, they had reached 97% accuracy on the word-level task. At a more fine-grained level, there were only clear significant changes in accuracy across trials in the very first run of training, in which performance in the very first Mini-Block was lower than the next three, but there were no significant differences between subsequent blocks. This suggests a fast-mapping effect, in which subjects already started learning after hearing each word only once (Carey & Bartlett, 1978; Heibeck & Markman, 1987). This same pattern of early learning was found in the Mini-Block analyses for RT. Additionally, there was an effect of Mini-Block for Runs 5 and 6, which most likely represents acclimation to performing the task in the scanner for the first time and to performing the task for the first time on Day 2, respectively. There was also a main effect of Mini-Block in Run 3 for both accuracy and RT, but this seems to be driven by unusually high performance on Mini-Block 3. It is unclear why this would be the case, but may just be an artifact of the small number of items in each Mini-Block, representing a Type I error.

As expected, learning did not happen as quickly or as completely at the sentence level as it did at the word level, but there was still clear evidence of learning. Performance continued to improve even in the last day of training, as there were no significant differences between accuracy for Run 2 and any of the runs in Day 2, but accuracy in Runs 7 and 8 was significantly
higher than accuracy in Run 2, demonstrating an improvement in performance in the last two runs of training. Furthermore, the only significant RT difference was between Runs 1 and 8. Mini-Block analyses were less clear at the sentence level, which may have to do with larger between-subjects differences in sentence learning as compared to word learning (see Appendix F). The fact that accuracy continued to increase and RT continued to decrease in Day 3, along with the fact that mean accuracy for participants maxed out around 80%, suggests that these learners, on average, did not reach high proficiency. This has major implications for the interpretation of fMRI analyses (see Section 4.3), as it cannot be claimed that this study examines the full course of learning, at least not for all subjects.

Performance on the GJT provided important insights as to why sentence-level learning was lower than anticipated. As indicated by $d^\prime$ scores, learners did significantly improve in their ability to detect syntactic violations, but they were essentially oblivious to morphosyntactic violations, at least in terms of response accuracy. However, by going beyond typical RT analyses of correct responses and analyzing RT according to response type (i.e., “Good” v. “Bad”), some evidence of morphosyntactic learning was revealed. Presumably, if learners were completely unable to distinguish morphosyntactic violations from matched grammatical sentences, they would treat these sentence types exactly the same, and therefore have similar RTs when endorsing grammatical and ungrammatical sentences, or when rejecting grammatical and ungrammatical sentences. There was no difference in RT between rejecting grammatical and ungrammatical sentences. However, on Day 3, after training was complete, learners were significantly faster at endorsing grammatical sentences than ungrammatical sentences with morphosyntactic violations. This suggests that learners may have hesitated before classifying
morphosyntactic violations as grammatical, possibly due to some subconscious rule knowledge. Still, few participants seemed to have accurate and complete knowledge of the morphosyntactic agreement rules of Mini-Basque (see Section 4.3).

Regarding production, performance improvements were seen over all three days for both words and sentences. For words, the number of correct words more than doubled between Days 1 and 2, and continued to improve in Day 3, but this performance (65.4% correct) was much lower than the word comprehension task, picture-matching. Clear improvements were also seen in the sentence production task. However, it should be noted that the nature of the word- and sentence-level production tasks require different coding schemes, and the sentence coding scheme focuses on elements of grammar rather than a holistic representation of the sentence, which results in a more lenient scoring system. Sentence production abilities were much poorer than comprehension abilities.

4.1.1 Research question 1. How does behavioral performance change over the course of L2 acquisition?

As predicted, accuracy increased and RT decreased as a function of exposure for both word- and sentence-level learning. Both also followed an asymptotic pattern, with the majority of learning-related performance changes being evident very early in the learning process and tapering off in later runs. In line with Hypothesis 1, these patterns were different for words and sentences. Words were learned very quickly, with the only significant performance increases occurring after just one exposure to each word. This is consistent with the literature on word learning, which has demonstrated that word learning is robust after very little exposure (Carey & Bartlett, 1978; Heibeck & Markman, 1987).
While learning at the sentence level was also most apparent early in the learning process with smaller gains later on, significant changes in performance were seen throughout the entire experiment. Unlike with words, where subjects reached a ceiling level of performance by the third run of exposure, sentence performance never surpassed 80% accuracy when averaged across subjects. This suggests that subjects still had more to learn about grammar, and would have continued to improve, given more exposure. Pilot testing demonstrated that learners could reach at least 90% accuracy in the same amount of training as the current experiment, and some subjects in this experiment reached 100% accuracy, so the mini-language should be learnable to high proficiency, given adequate exposure, and perhaps ideal cognitive resources. It is also possible that participants would have performed better had they not been performing the task in the MRI environment (Karuza, Li, Weiss, & Aslin, 2012).

Within words, there seemed to be an advantage for adjectives over both nouns and verbs in accuracy and RT, and an advantage of nouns over verbs in RT only. The latter is consistent with previous findings showing that L2 nouns are easier to learn than verbs (Getner, 1982; Gillette, Gleitman, Gleitman, & Lederer, 1999; Kallkvist, 1999; Lennon, 1996; Mestres-Missé et al., 2010). However, these differences could be due to other factors, such as word length (nouns = 2.5 syllables, adjectives = 2.75 syllables, verbs = 3.25 syllables), or potentially higher saliency for colors than animal names or verbs.

As expected, differences were also observed between sentence conditions. Somewhat surprisingly, there were no significant differences between word order and lexical trials, suggesting that both are learned early in the learning process. In line with predictions and previous research (e.g., DeKeyser, 2005; Johnson & Newport, 1989; Newport, 1990), accuracy
for the agreement condition was significantly lower than the other conditions in the early stages of learning, suggesting that these rules were initially harder to learn. However, performance in this condition eventually caught up to the other two. Results from the GJT also demonstrate differences between knowledge about word order and agreement, with substantially higher performance on syntactic violations than on morphosyntactic violations, for which performance was incredibly poor. In fact, learners were more likely to endorse morphosyntactic violations than to reject them which, along with $d'$ scores, suggests an inability to distinguish these violations from grammatical sentences. However, RT analyses suggest that there may be some unconscious knowledge of morphosyntactic agreement, since learners were slower to endorse morphosyntactic violations than grammatical sentences. While this pattern of results is compatible with Hypothesis 1, it was expected that learners would show at least some improvements in accuracy in the domain of morphosyntax. This issue is discussed in detail in Section 4.3.

Finally, Hypothesis 1 was further confirmed by results from the production task, in that there were clear improvements in this skill over time, but performance was much lower than in the picture-matching tasks, which assessed comprehension.

4.2 fMRI Results

Research questions 2 through 5 are addressed by results from the fMRI analyses.

4.2.1 Research question 2. Are the neural correlates of adult L2 acquisition the same as the neural correlates of L1 processing?

A large network of brain regions were activated during learning and processing of Mini-Basque. For word-level learning, the only region that was significantly activated in whole-brain
analyses was the parahippocampal gyrus, which has been found to play a role in lexical/semantic learning and processing in L2 (Breitenstein et al., 2005; Mestres-Missé et al., 2008, 2010; Raboyeau et al., 2004). Lexical/semantic processing at the sentence level resulted in deactivation in bilateral STG, which is in contrast to previous findings showing activation in this area for lexical/semantic processing in both L1 (Friederici et al., 2003; Grey et al., 2013; Illes et al., 1999; Kuperberg et al., 2000; Newman et al., 2001b; Tagarelli et al., 2012; Ullman et al., in prep) and L2 (Newman-Norlund et al., 2006) lexical/semantic processing. The left inferior parietal lobe was activated during the lexical condition, which is consistent with findings from our ALE meta-analyses of the functional neuroanatomy of L1 and L2 (Grey et al., 2013; Tagarelli et al., 2012; Ullman et al., in prep), as well as studies of word learning (Breitenstein et al., 2005; Cornelissen et al., 2004; Mestres-Missé et al., 2008, 2010). Several other areas also activated during the lexical conditions are not typically activated during lexical/semantic processing or learning, but have been shown to be involved in learning in word segmentation and artificial grammar learning studies. These include the cerebellum (Fletcher, Büchel, Josephs, Friston, & Dolan, 1999; Mestres-Missé et al., 2010), inferior occipital lobe (Forkstam et al., 2006; Mestres-Missé et al., 2010; Petersson et al., 2012, 2004), supramarginal gyrus (Karuza et al., 2013; McNealy, 2006; Scott-Van Zeeland et al., 2010), and putamen (Karuza et al., 2013; McNealy, 2006; Mestres-Missé et al., 2008; Scott-Van Zeeland et al., 2010). Additionally, areas typically involved in L1 and grammar processing, specifically left BA44 and BA6, were significantly activated for the lexical condition. The middle frontal gyrus was also activated in this condition, which is frequently found to be involved in language learning across multiple
For general sentence-level learning, several regions that have been shown to be involved in L1 grammar processing were activated, included BA44 and BA6 (Dapretto & Bookheimer, 1999; Friederici et al., 2003; Grey et al., 2013; Newman et al., 2001a; Tagarelli et al., 2012; Ullman et al., in prep), as well as areas that may be involved in various stages of L2 grammar processing, like BA45 (Golestani et al., 2006; Grey et al., 2013; Hasegawa et al., 2002; Nakai et al., 1999; Rüschemeyer et al., 2005; Tagarelli et al., 2012; Ullman et al., in prep; Wartenburger et al., 2003), and L2 grammar learning, including the inferior parietal lobe (McNealy, 2006; Scott-Van Zeeland et al., 2010) and the middle occipital lobe (Fletcher et al., 1999; Hauser et al., 2012). For the feedback contrasts, in addition to the IFG, bilateral putamen was activated in ROI analyses, supporting a role for the basal ganglia in L2 grammar learning (Abutalebi, 2008; Karuza et al., 2013; McNealy, 2006; Newman-Norlund et al., 2006; Petersson et al., 2012; Scott-Van Zeeland et al., 2010; Ullman, 2004). As in the lexical condition, both the word-order and agreement conditions showed activation in the inferior parietal lobe and the right putamen, and deactivation in the STG. The only other area activated in the agreement condition was BA47. The word order condition activated a larger network of regions, some of which overlapped with the lexical condition, including the cerebellum, the inferior occipital lobe, the supramarginal gyrus, and the middle frontal gyrus. Additional activation for the word order condition was found in the insula, middle occipital lobe, superior frontal gyrus, and thalamus.

In the GJT, which provides a measure of L2 grammar processing, syntactic violations activated several regions in the IFG, including BA44, BA45, BA47, the precentral gyrus (BA6),
and the caudate and putamen, though the latter three were not significant. This pattern of activation suggests that L2 grammar processing depends on regions that have been shown to be involved in L1 grammar processing, like BA44, BA6, and BA45 (Dapretto & Bookheimer, 1999; Friederici et al., 2003; Grey et al., 2013; Newman et al., 2001a; Tagarelli et al., 2012; Ullman et al., in prep), areas that are involved in L1 lexical/semantic processing, like BA45 and BA47 (Grey et al., 2013; Tagarelli et al., 2012; Ullman et al., in prep), as well as areas that seem to subserve L2 grammar learning, like the basal ganglia (Abutalebi, 2008; Grey et al., 2013; Karuza et al., 2013; McNealy, 2006; Newman-Norlund et al., 2006; Petersson et al., 2012; Scott-Van Zeeland et al., 2010; Tagarelli et al., 2012; Ullman, 2004). Morphosyntactic violations did not activate any regions significantly.

While many of the same areas thought to be involved in L1 processing were activated in the Mini-Basque learners, especially in the IFG, the laterality of this activation was not always typical of L1. In L1, activation is generally found in the left IFG, and other regions in the left hemisphere of the brain (Dapretto & Bookheimer, 1999; Friederici et al., 2003; Grey et al., 2013; Illes et al., 1999; Newman et al., 2001a; Poldrack et al., 1999; Tagarelli et al., 2012; Ullman et al., in prep). However, in the Mini-Basque learners, activation was often bilateral, and frequently right-hemisphere dominant. For example, in sentence-level learning, only areas in the right IFG were significantly activated. This finding is consistent with studies on L2 processing, where activation tends to be more bilateral, and may be right-hemisphere dominant for some structures, as demonstrated in Figure 30 from a meta-analysis on the functional neuroanatomy of L2 (Grey et al., 2013; Tagarelli et al., 2012; Ullman et al., in prep). More extensive activation in L2 has been attributed to more effortful processing in a weaker language (Abutalebi, 2008; Perani &
Abutalebi, 2005), but the reason for bilateral activation and right-hemisphere dominance remains unclear.

Figure 30. Activation-likelihood estimation meta-analysis of L1 (red) and L2 (blue) grammar activation (pink indicates and overlap of L1 and L2 activation). L2 activation is more extensive and bilateral, with right-hemisphere dominance for some structures. The meta-analysis includes data from 16 groups of subjects in eight empirical papers (Ullman et al., in prep). y = 13.

These findings demonstrate that for both lexical/semantics and grammar, at both the word and sentence levels, L2 acquisition and processing involve some areas that are activated in L1 processing, as well as some additional areas that are not typically activated in L1. These findings contradict the Fundamental Difference Hypothesis, which predicts no overlap between the neural bases of L1 and L2. They support both the Convergence Hypothesis and the DP model, especially because several of the regions for which these models predict involvement are indeed activated during the learning and processing of Mini-Basque, including the IFG, the basal
ganglia, the hippocampus and medial temporal regions, temporal cortex, and parietal cortex. The ways in which these patterns of activation change over time, particularly as a function of language domain, should help to further distinguish these two theories.

4.2.2 Research questions 3 and 4. Does neural activity change over the course of adult L2 acquisition? If so, how? Do changes in neural activity and behavioral performance differ according to aspects of language, specifically lexical/semantics, syntax, and morphosyntax?

Research question 2 examined which parts of the brain were involved during language tasks in an L2, and whether these differ from those areas involved in L1 processing. Research questions 3 and 4 should be answered together, as they get at the overarching issue of how to characterize the neural changes during learning for different language domains. This section will revisit results from each of the fMRI tasks as a function of time, and then summarize the differences and overlap in the neural regions recruited for the various language domains.

During word learning, there was no significant activation in Day 1. However, some interesting trends emerged regarding activation during the word learning task. There was marginally significant activation in the bilateral hippocampus on Day 1, and significant activation in the right parahippocampal gyrus in Day 3, which is consistent with predictions of the DP model and previous findings that have suggested a role for the hippocampus and MTL in forming new semantic representations (Breitenstein et al., 2005; Mestres-Missé et al., 2008, 2010; Raboyeau et al., 2004; Ullman, 2004). In addition, there was deactivation in areas of the IFG, including BA44, BA45, and BA6, though the latter was not significant. Interestingly, there was a trend toward activation of BA44 and BA45 on Day 3, suggesting a shift in the involvement
of these regions as learning progressed. Finally, though again only marginally significant, activation in the right putamen was apparent on Day 3 only.

During sentence learning, only right BA45 was activated on Day 1, and right prefrontal cortex (BA9) was deactivated. On Day 3, activation in right BA45 was maintained, but extended to right BA44 and BA6. This finding is of particular interest because it demonstrates an increase in IFG activation over the course of learning, with BA45 involved in early stages and BA44 and BA6 being recruited in later stages of learning. Such a pattern is consistent with the DP model (Ullman, 2001c, 2004, 2005), and contradicts the Convergence Hypothesis, which predicts greater involvement of the IFG in early stages of learning, with decreasing involvement as L2 proficiency increases and the effort involved in L2 processes consequently decreases (Abutalebi, 2008; Perani & Abutalebi, 2005). However, this activation is mostly right-lateralized, which is in contrast to the left-lateralized predictions of these models. The middle occipital lobe (BA19), which has repeatedly been shown to be involved in language learning (Fletcher et al., 1999; Forkstam et al., 2006; Hauser et al., 2012; Petersson et al., 2012, 2004), was also activated on Day 3. Deactivation in the right prefrontal cortex was maintained between Day 1 and Day 3, with additional deactivation in bilateral anterior cingulate cortex (Fletcher et al., 1999; McNealy, 2006; Mestres-Missé et al., 2008, 2010; but see Newman-Norlund et al., 2006), left insula (Mestres-Missé et al., 2008), and left supramarginal gyrus, which is inconsistent with previous research (Karuza et al., 2013; McNealy, 2006; Scott-Van Zeeland et al., 2010).

During the GJT, syntactic violations activated bilateral BA47 on Day 1, though this was only marginally significant in the left hemisphere. BA6 was deactivated on Day 1. On Day 3, left BA44 and right BA47 were significantly activated, and there was marginally significant
activation in bilateral BA45. This again demonstrates an increase in involvement of the IFG as proficiency increases, and again shows a role for BA44 in later stages of L2 grammatical processing. Furthermore, in a contrast between syntactic violations and grammatical sentences where Day 1 was subtracted from Day 3, activation was found in the left caudate and right putamen. This pattern of increased basal ganglia activity over time is again consistent with the DP model (Ullman, 2001c, 2004, 2005), and contradicts the Convergence Hypothesis (Abutalebi, 2008; Perani & Abutalebi, 2005). Interestingly, the right putamen was also found to be involved in learning-related changes in the word-level picture-matching task, as well as for the three sentence conditions on the feedback event. No changes in activation were observed over the course of learning for morphosyntactic violations, but this is likely due to the lack of substantial learning of this particular aspect of Mini-Basque.

4.2.3 Research question 5. Is there a relationship between behavioral and neural aspects of L2 learning? If so, does this change over the course of learning?

Research questions 3 and 4 examined the relationship between exposure and neural activity over the course of L2 learning. As demonstrated in Chapter 3, exposure is not directly related to learning, and so the neural bases of L2 learning cannot be thoroughly examined through observations of neural changes over time. While exposure is undoubtedly an important contributor to learning, understanding the relationship between proficiency and neural activity can add a crucial piece to this puzzle. To do this, the relationship between behavioral performance and neural activation was tested.

At the word level, there were no significant correlations with any ROIs when corrected for multiple comparisons. However, there were several trends toward positive correlations
between performance and activation in the left hippocampus, the left parahippocampal gyrus, left BA44, and bilateral BA47. In other words, learners who performed better on the word learning task in the first run had more activation in these areas. The pattern for hippocampal activation is similar to findings that better learners show less hippocampal suppression over time during word learning (Breitenstein et al., 2005). As BA47 has been implicated in word learning, particularly retrieval (Dapretto & Bookheimer, 1999; Illes et al., 1999; Mestres-Missé et al., 2008; Poldrack et al., 1999; Ullman, 2004), it is not surprising that learners who perform better would recruit this area. The relationship between BA44 and performance is more puzzling, in light of previous research.

The relationship between behavioral performance and hippocampal and MTL activation at the sentence level was quite revealing regarding the role of these structures in learning. Activity in the bilateral parahippocampal gyrus and the left hippocampus were significantly negatively correlated with behavioral performance on the sentence-level picture-matching task in the very first run of training. This pattern was also observed for the right hippocampus, but not significantly. This pattern persisted in Run 3, but the correlation coefficients decreased and the relationships were no longer significant. No relationship between performance and hippocampal activation was observed after Run 3. In Run 5, about halfway through training, there was a strong (but not significant) negative correlation between behavioral performance and activity in left BA45. There was a positive correlation between performance and activation in left BA6. Within the IFG, there were also marginally significant positive correlations between accuracy and activation in right BA6 and left BA44.
The patterns above suggest that in early stages of grammar learning, the hippocampus and MTL are activated, and learners who performed poorly on the sentence comprehension task relied more on these regions. As learning progresses, there is no longer a difference between good and poor learners with respect to hippocampal and MTL involvement. After over an hour of grammar training and overnight consolidation, a neocortical area typically associated with retrieval in lexical/semantics (BA45; Dapretto & Bookheimer, 1999; Illes et al., 1999; Mestres-Missé et al., 2008; Poldrack et al., 1999; Ullman, 2004) seems to take the place of the hippocampus and MTL, with poorer learners relying more on this structure than better learners. The better learners, on the other hand, rely more on neocortical structures typically involved in L1 grammar processing, namely BA44 and BA6. These findings demonstrate a shift from hippocampal and MTL structures in early learning to BA45 as lexical memories are consolidated, and eventually to BA44 and BA6 as grammar proficiency increases in conjunction with pattern learning.

For the GJT, the only significant correlation between performance and activation was found in right BA47 on Day 1, suggesting a role for processes typically involved in lexical/semantic processing in the early stages of learning (i.e., after only about 30 minutes of sentence-level training). On Day 3, there were marginally significant positive correlations between performance and activation the IFG, including BA44, BA45, BA47, and BA6. There were also marginally significant positive correlations between performance and activation in the basal ganglia, including the caudate and putamen. While these correlations are not significant, they suggest a trend toward a shift in reliance from areas involved in processing arbitrary information to areas involved in pattern recognition and grammar processing.
4.2.4 Implications for Neurocognitive Theories of L2 Acquisition and Processing.

The findings from the current study have important implications for teasing apart the three broad frameworks of L2 neurocognition. While the significant results are meaningful on their own, there are clear trends that repeat themselves over and over again, and these should also be accounted for, particularly because many of them involve marginally significant correlations with moderate to large effect sizes (Cumming, 2014; Kline, 2004; Norris, 2013).

The first main distinction between the Fundamental Difference Hypothesis, the Convergence Hypothesis, and the DP model involves their predictions for overlap in neural activity for L1 and L2. To review, the Fundamental Difference Hypothesis predicts that there should be no overlap between the neural structures underlying L1 and adult-learned L2 (Bley-Vroman, 1989), the Convergence Hypothesis predicts the neural correlates of L1 and L2 are largely overlapping (Abutalebi, 2008), and the DP model predicts overlap in some areas but not in others, as a function of language domain and proficiency (Ullman, 2004). In the current study, activation during L2 learning and processing was found in some areas that are typically activated in L1, like BA45 and 47 for lexical/semantics and BA44 and 6 for grammar, as well as in areas that are not typically activated in L1, like the hippocampus, MTL, and basal ganglia, and some areas of the prefrontal cortex for grammar. Additionally, right-hemisphere homologues of IFG structures were also activated, which is not typical of L1 processing. These findings therefore contradict the Fundamental Difference Hypothesis, and are consistent with the DP model. Although the Convergence Hypothesis mainly predicts overlap between L1 and L2 activation, it also predicts additional activation in the L2 in early stages of learning, particularly in the prefrontal cortex and basal ganglia, because of the effortful processing involved in L2 use
(Abutalebi, 2008; Indefrey, 2006). Therefore, it is not possible to distinguish the two latter models based on this information alone. However, each theory makes different predictions regarding changes in activation over the course of learning, so further analyses can tease them apart.

At the most basic level, the Convergence Hypothesis predicts a decrease in activation in the IFG and basal ganglia as proficiency increases (Abutalebi, 2008), whereas the DP predicts increased activation, in BA45 and 47 for lexical/semantics, and in BA44 and BA6 for grammar (Ullman, 2004). More specifically, the DP model predicts a shift from hippocampal and MTL structures in early lexical/semantic learning to neocortical regions (BA45 and 47 in the IFG, as well as the temporal lobe) as learning progresses. For grammar, the DP model also predicts this hippocampal/MTL to neocortical shift, but also predicts involvement of the basal ganglia as grammar learning shifts to the procedural memory system, and eventually to BA44 and 6 as grammatical representations are solidified (Ullman, 2004).

To assess the validity of these predictions, I first turn to lexical/semantics. At the word level, the only region activated in a whole brain analysis was in the parahippocampal gyrus. Upon closer inspection, there was evidence of hippocampal activation on Day 1 but not in Day 3, suggesting a decrease in hippocampal involvement over time, though parahippocampal activation remained significant in Day 3. Activity in the hippocampus and parahippocampal gyrus were also positively correlated with performance on the word-level task, suggesting that better learners were more able to take advantage of these neural resources when learning new vocabulary. Regarding the IFG, there was deactivation in this region on Day 1, and activation on Day 3, suggesting an increase in recruitment of this area over the course of lexical/semantic
learning, and a shift from hippocampal to neocortical regions, consistent with the DP model. Within the IFG, these patterns were found for BA45, which is predicted by the DP model, but also for BA44, which is somewhat surprising, given that this area is generally thought to be involved in (morpho)syntactic processing. This may be an effect of performing a word-level task on words that have been trained in a sentence-level context. However, regardless of the predictions of these neurocognitive models, BA44 activation in L2 lexical/semantic processing is consistent with previous findings (Abutalebi et al., 2013; Chee et al., 2001; De Bleser et al., 2003; Grey et al., 2013; Liu, Hu, Guo, & Peng, 2010; Luke et al., 2002; Stein et al., 2009; Tagarelli et al., 2012; Tatsuno & Sakai, 2005; Ullman et al., in prep; Vingerhoets et al., 2003; Wartenburger et al., 2003; Yokoyama et al., 2006), even if it is not predicted by any models.

The patterns of activation changes observed at the sentence level lend further support to the DP model. In the early stages of learning, BA45 was activated during grammar learning, and this activation was maintained throughout the end of training, suggesting a sustained role for declarative memory in grammar learning, at least over a short training period. By Day 3, BA44 and BA6 were also activated during grammar learning. A similar shift in IFG activation was observed in the GJT. This supports Ullman’s (2001a, 2004, 2005, 2012) claim that adult learners can proceduralize an L2 grammar, while still relying to some extent on declarative memory (BA45), in this case after only a few hours of training. It also demonstrates an increase in neocortical activation over the course of learning, which contradicts the Convergence Hypothesis (Abutalebi, 2008).

Correlational analyses between behavioral performance and neural activation delve even deeper into characterizing the neural underpinnings of adult L2 acquisition. At the very
beginning of sentence-level training, there was a negative correlation between performance and hippocampal/MTL activation. This is the inverse of the relationship seen at the word level, where better learners exhibited more activation in these regions, possibly because they were better at forming representations of new words in declarative memory. The relationship between performance and hippocampal/MTL activation at the sentence level may indicate that poorer learners rely more on declarative memory structures in early phases of grammar learning, whereas better learners are not relying as much on these structures, perhaps because they have begun to recruit redundant but more efficient procedural memory structures. This correlation decreases to the point of being non-significant by the beginning of Day 2, and disappears completely by the second run of Day 2. This change may suggest that as learners are exposed to more and more of an L2, even the poorest learners start to shift away from declarative learning mechanisms involved in learning and consolidation. Interestingly, about halfway through training, there was a non-significant but strong negative correlation between performance and activation in BA45, which may suggest that later on in learning, poorer learners are still depending more on declarative memory than better learners, but by this point they are recruiting retrieval mechanisms in this memory system, which may be different from those associated with BA47, which was also negatively correlated with performance in Run 1. Halfway through training, there were suggestive positive correlations between performance and activation in BA6 and BA44. It is possible, then, that while poorer learners rely more heavily on declarative memory structures, better learners have begun to automatize grammar skills and are therefore recruiting procedural memory structures.
Also consistent with the DP model (or the Convergence Hypothesis, for that matter) is the finding of basal ganglia activation, particularly the left caudate and right putamen. The caudate is predicted to play an important role in skill acquisition (Ashby et al., 2010; Doyon et al., 2009), and therefore grammar learning, but perhaps not in the very early stages of grammar (Ullman, 2004). Activation in the left caudate increased from Day 1 to Day 3 for syntactic violations. There was also a marginally significant correlation between the change in performance on GJT and activation in the caudate. The putamen, especially in the right hemisphere, also seemed to have a role in learning, for both lexical/semantics and grammar. Beyond being involved in the learning of language patterns (Karuza et al., 2013; Scott-Van Zeeland et al., 2010), the basal ganglia have been shown to be involved in general learning, reinforcement, and response selection (Desmond, Gabrieli, & Glover, 1998), the learning of arbitrary associations (Laubach, 2005), which is an important aspect of vocabulary acquisition, and meaning acquisition in language (Mestres-Missé et al., 2008). In the word-level task, there was marginally significant activation in the right putamen in Day 3, but not Day 1. For the GJT, when activation in Day 1 was subtracted from Day 3, activation in the right putamen was found. Additionally, there was a marginally significant correlation between performance increases and activation in the left putamen for the GJT. The right putamen was also activated for all sentence-level conditions – lexical, word order, and agreement – when analyses were collapsed across all runs. Taken together, these findings suggest that the putamen plays a role in L2 learning and processing, for both lexical/semantics and grammar, and may be more involved at later stages of learning, when representations have been consolidated. However, it should be noted that there was deactivation in the putamen for the main sentence-level task in an ROI analysis collapsed
across all runs, which contradicts the previous statement. Regardless, it seems quite possible that the putamen has an interesting role to play in L2 acquisition, and further analyses and research may shed more light on this structure.

4.2.5 Implications for Second Language Acquisition Research. The results from this study also have implications for research in SLA and applied linguistics. While the field of SLA has made great strides in the past 40 years on so in characterizing L2 learning and processing, many of the fundamental questions that SLA researchers ask cannot be answered with behavioral methods alone. In particular, this study provides insights into the nature of age effects in language learning, the relationship between L1 and L2 processing in the brain, the extent to which L2 learning and processing rely on implicit or explicit mechanisms, and the operationalization of L2 proficiency.

Because L1s are generally learned early in life, aside from relatively rare cases, and the focus of this dissertation is on late-learned L2, the issues of age effects and the relationship between L1 and L2 are necessarily intertwined here. FMRI activation in the Mini-Basque learners demonstrated the involvement of neocortical areas typically activated in L1 processing, like the inferior parietal lobe and LIFG, particularly BA44, and BA47 for lexical/semantics, and BA44 and BA6 for grammar. The fact that these areas are activated in the Mini-Basque learners demonstrates that several of the regions that are involved in the processing of an early-learned L1 are still available to, and recruited by, late-L2 learners. In other words, as individuals get older and learn a new language, they do not completely lose access to the neural structures that are involved in L1. In spite of this overlap, however, we cannot be sure that L2 learners are using these mechanisms in the same way that L1 speakers do. This remains an empirical question.
Furthermore, in addition to these L1-language areas, there were also regions activated in the Mini-Basque learners that are *not* typically activated in L1 processing, which suggests that the L1 language regions, while engaged, are not sufficient for L2 learning and processing. First, activation was observed in regions that are thought to be involved in learning and memory consolidation, like the hippocampus, MTL, basal ganglia, and possibly the middle occipital lobe. It makes sense that these areas would be involved in an L2 learning paradigm, but not in L1, since L1 processing studies are generally performed on adults who have already learned their language (but see Skeide et al., 2014). Second, while L1 activation is typically found in the left hemisphere, Mini-Basque learners demonstrated more extensive and bilateral activation, and actually showed right hemisphere dominance for many structures. Greater activation in L2 relative to L1 is likely due to the increased effort involved in learning and processing an L2 (Abutalebi, 2008; Golestani et al., 2006; Hernandez & Meschyan, 2006; Meschyan & Hernandez, 2006), but the reason for right hemisphere dominance is still unclear. Taken together, this suggests that there is not a critical period for the involvement of certain neural structures in late-learned L2 (though it remains to be seen whether there is a critical period for the function of these structures), but that in adulthood, L1 neural mechanisms are not sufficient for L2 learning and processing, and other structures and neural systems must be recruited.

The activation of certain neural structures in the Mini-Basque learners, particularly those of the procedural memory system, may help clarify a key issue in SLA research regarding the development of implicit and explicit knowledge, the points at which learning and knowledge might become implicit, if ever, and the relationship between the two. The procedural memory system is considered to be one of several implicit memory systems in the brain, so the structures
within this system, especially the basal ganglia, are expected to be involved in implicit processes (Packard & Knowlton, 2002). If L2 learning and processing is always explicit, then these structures should never be activated. However, Mini-Basque learners showed activation in the caudate nucleus and putamen (structures within the basal ganglia), as well as BA44 and BA6, which suggests that at least some L2 learning or processing can be implicit. Additionally, early reliance on the hippocampus and MTL and later recruitment of the basal ganglia suggest that learning and processing in L2 may not initially depend on implicit mechanisms, but may shift to those processes after a sufficient amount of exposure or level of proficiency. Of note, however, is that some declarative memory structures, especially BA45, are activated in later stages of Mini-Basque learning. This suggests that both implicit and explicit processes may be involved simultaneously in L2. Such redundancy is not uncommon among biological systems (Tononi, Sporns, & Edelman, 1999). As with the involvement of L1 structures in L2, it should be noted that there may be alternate explanations for the involvement of the basal ganglia in L2 learning and processing that do not involve implicit mechanisms (see Abutalebi, 2008). This is, again, an empirical question.

Finally, the results from neuroimaging studies like this one may help with the issue of operationalize proficiency. The field of SLA is rife with constructs that are historically difficult to operationalize (Gregg, 1984; B. McLaughlin, 1978), and proficiency is arguably the most elusive term to define (Larsen-Freeman & Long, 1991). Recent neuroimaging research has demonstrated that behavioral evidence alone might not be sufficient for determining proficiency, since different groups of speakers may perform similarly on language tests, but demonstrate different brain patterns (Morgan-Short, Steinhauer, et al., 2012; Mueller, 2005; Mueller et al.,
2005). Regarding learners who exhibit different levels of behavioral performance, findings from this dissertation show that those learners who perform poorly on L2 tasks rely on different structures over the course of learning from those who perform well. In particular, for grammar, poorer learners show a greater reliance on structures associated with more declarative memory, like the hippocampus and MTL, whereas better learners demonstrate a shift to structures associated with procedural memory and implicit processes, like the basal ganglia, BA44, and BA6. This suggests that the profile of a highly proficient learner might go beyond high performance on behavioral tasks and include activation of certain neural regions that suggest the ability to automatize grammar. The goal of L2 teaching, then, might not only be to increase scores on language tests, but also to determine what sorts of training conditions and cognitive abilities can facilitate the involvement of these automatic processes in the brain.

4.3 Limitations and Future Directions

While this study has hopefully made significant contributions to the fields of SLA, cognitive psychology, and cognitive neuroscience, it is certainly not without limitations. In this section, I outline some of those limitations.

The first limitation of this study relates to the ecological validity of using a mini-language model. Although I have outlined the strengths of using a mini-language above (see Section 1.2.2.1.5), there are still limitations associated with this approach. In this dissertation, I have attempted to make the case for studying a mini-language in lieu of a full natural language, based on two fundamental issues. The first is that while our goal is to understand the neural and behavioral underpinnings of actual language learning in the real world, practical concerns make it impossible to do so with fine-grained tools and measurements. Natural L2 learning research
necessarily misses aspects of the learning trajectory because it is impossible to follow learners over the course of learning an entire language. For this reason, artificial linguistic systems can be extremely useful models for investigating language learning in a longitudinal and continuous manner. However, artificial linguistic systems are devoid of many of the features of natural language, and therefore findings from studies using these systems may not adequately generalize to natural languages. Mini-languages allow for the longitudinal, continuous investigation of L2 learning, while maintaining many of the features of natural languages, certainly more so than artificial linguistic systems. Nevertheless, a mini-language is still a model.

The second limitation of this study is also related to ecological validity, particularly that of the learning paradigm and setting. In this study, learners were exposed to the mini-language through a series of picture-matching tasks, with intermittent comprehension and production tasks, and little interaction. This is very different from natural language learning settings in either immersion or classroom situations. Furthermore, learners completed many of the language learning tasks lying supine in an MRI scanner, which is generally very loud and uncomfortable. These first two limitations represent compromises that must be made in order to test specific questions about language learning and the brain in a controlled environment, but caution should be taken when generalizing results from this study to the real system of interest, that is, natural languages.

The third limitation of this study is the number of subjects, which limits the power for statistical analyses. The fMRI analyses in this study were conducted on 13 or 14 subjects (see Section 2.1). While this is a reasonable number of subjects for an fMRI study (Abutalebi et al., 2013; Chee et al., 2001; De Bleser et al., 2003; Ding et al., 2003; Grey et al., 2013; Halsband et
al., 2002; Hernandez et al., 2007; Hernandez & Meschyan, 2006; Luke et al., 2002, 2002; Meschyan & Hernandez, 2006; Rüschemeyer et al., 2005; Sakai et al., 2004; Saur et al., 2009; Stein et al., 2009; Tagarelli et al., 2012; Tatsuno & Sakai, 2005; Vingerhoets et al., 2003; Wartenburger et al., 2003; Yokoyama et al., 2006, 2009), especially considering the cost and time involved in a study of this length, this does not eliminate the issues related to power with such a small number of subjects. Furthermore, FWE corrections in fMRI analyses result in rather strict statistical thresholds, as evidenced by the limited number of contrasts that resulted in significant activation in whole-brain analyses. Many of the results reported in this dissertation are marginally significant trends, which were included in spite of the lack of significance because the trends were robust across multiple analyses, and often were the results of correlations with moderate to large effect sizes, which may be more important than $\alpha$-levels when interpreting relationships between variables (Field, 2012). Indeed, recent work in SLA cautions against over-interpreting significance levels at the expense of noticing actual patterns in the data (Cumming, 2014; Kline, 2004; Norris, 2013). In any case, results reported on marginally significant trends should be interpreted with caution. Further measures will be taken to determine whether these findings are indeed robust, including running Monte Carlo simulations to determine an appropriate cluster-size cutoff for uncorrected whole-brain analyses (Ward, 2000).

The fourth limitation of this study is the fact that most subjects did not reach “high proficiency” in grammar competence, as evidenced by their performance on the sentence-level picture-matching and grammaticality judgment tasks. The morphological features in Mini-Basque, especially the determiners, are redundant and opaque, which make them especially difficult to learn (DeKeyser, 2005). They are redundant because every sentence contained
multiple cues regarding number: (1) the determiners on the subject and object, and (2) the auxiliary at the end of the sentence. Every combination of determiners on the subject and object matched a specific auxiliary, so the subject could either know that –ek on the subject and –ak on the object meant “plural subject, plural object,” or that the auxiliary, dituzte, indicated the same thing. The auxiliary seemed to be more salient to the subjects in this study (based on results not discussed here from sentence production and questionnaires), which is most likely because they were easier to hear, relatively distinct from one another phonologically, and came at the end of the sentence. The determiners, on the other hand, were harder to hear and did not differ greatly from one another phonologically. Furthermore, they were opaque in that the –k on the end of a determiner could either indicate plurality, ergativity, or both. Finally, the determiner –ak was ambiguous in that it could denote a singular ergative subject, or a plural absolutive object. Regardless of redundancy and opaqueness, agreement patterns and verb conjugations tend to be particularly difficult for L2 learners (Brooks et al., 2006), so the fact that the Mini-Basque learners had the most trouble with these features of the language is not surprising.

Although performance was similar across conditions by the end of training in the sentence-level picture-matching task, performance was incredibly poor on the morphosyntactic violation condition in the GJT, even in Day 3. In retrospect, this is likely due to a design issue in the sentence-level picture-matching task, which was the main training task for Mini-Basque grammar. In the agreement trials for this task, subjects heard a sentence and had to choose the picture that had the correct number of animals in the subject and object position. Because of the redundancy between the determiners and the auxiliary in Mini-Basque morphology (described just above), it was possible to choose the correct picture by paying attention to either the
determiners or the auxiliary, but not necessarily to both. It seems likely that subjects were not attending to the actual agreement between the determiners and the auxiliary, since it was aspects of the morphology (i.e., determiners or auxiliary), but not morphosyntactic agreement (i.e., the relationship between the determiners and the auxiliary), that was task essential for this condition (Loschky & Bley-Vroman, 1993). In other words, in order to complete the task on agreement trials, learners had to know something about determiners or the auxiliary in the sentence, but not necessarily how they worked together. This likely posed an obstacle to learning Mini-Basque grammar to high proficiency.

The inability of subjects to reach high proficiency by the end of Mini-Basque training has major implications for the interpretation of results from this study. Crucially, contrary to its intended purpose, this study does not examine the full course of learning for all subjects. This raises the question of whether subjects would have reached high proficiency with more exposure to the mini-language, or with different instructional conditions (e.g., a more explicit training condition). One possibility is that learners might be forced to process the mini-language at a higher level if they received feedback on the production task, as was the case in the BROCAN TO2 studies (Grey, 2013; Morgan-Short et al., 2010; Morgan-Short, Steinhauer, et al., 2012). On the other hand, it is possible that the complex morphosyntax of Mini-Basque is too difficult for L2 learners. However, this latter explanation is unlikely, as large individual differences between subjects were observed, with some subjects reaching high proficiency, as operationalized by very high accuracy (see Appendix F). This study does provide a continuous examination of language learning, which did indeed involve changes in brain activation. By capitalizing on individual differences between subjects and running further analyses on high-
performing learners, patterns of brain activation involved in continuous learning to high proficiency can be revealed.

The fifth limitation of this study is not a limitation, *per se*, but rather a consideration that should be made when interpreting results from fMRI studies. Within fMRI studies of L2 learning and processing, there is a variety of tasks that have been employed to study L2 neural activation (e.g., picture naming, GJTs, sentence completion, verb matching), as well as a variety of baselines, or control conditions, against which activation from these tasks are compared. Some baselines, like the one used in this study, are designed to subtract out any task-related activation not of interest to the main question (in this case, L2-related activation). Other baselines, like fixation, only subtract out resting state activity, and therefore allow a lot of task-related activation not related to L2 (e.g., visual, auditory, and motor processing) to show up in the results. As a result of these different tasks and different baselines, there are slightly different findings across studies. However, in spite of this heterogeneity in experimental design, there is still a substantial amount of convergence across studies, which is demonstrated in meta-analyses by the large number of studies that contribute to clusters of activation in typical language and learning areas in L2 (Grey et al., 2013; Indefrey, 2006; Stowe & Sabourin, 2005; Tagarelli et al., 2012; Ullman et al., in prep). These considerations are important not only for interpreting fMRI results, but for interpreting any results from a set of empirical studies claiming to examine similar constructs with different experimental designs, groups of subjects, and control tasks or groups.

The sixth limitation of this study is a more theoretical one pertaining to the issue of whether the neural bases of *learning* and *processing* can be distinguished. As mentioned in the
review of literature (see Section 1.2), most language studies focus on processing rather than learning. Even studies that claim to investigate learning in longitudinal designs are really looking at processing at two different points in time. It is very difficult to design a task that investigates learning, *per se*, and it may be impossible to learn without simultaneously processing language. One potential solution to this problem is to determine whether the brain regions activated in a particular task are regions that are known to be involved in learning. For example, the hippocampus and basal ganglia were both activated over the course of the Mini-Basque experiment, which suggests that the fMRI tasks captured learning. One might also think of learning as processing at a more fine-grained level. Perhaps by observing more and more frequent changes in processing – from months, to days, to minutes, to trials – we can characterize the mechanisms involved in learning. Another method that has been used to investigate the neural bases of learning is delta analyses, which investigate the extent to which neural activation is related to changes in performance (e.g., Karuza et al., 2013). Future analyses on the Mini-Basque learners will employ delta analyses. Whether learning and processing can truly be distinguished empirically, and how, is an interesting question for future research.

### 4.4 Future Directions and Conclusions

The work in this dissertation has laid the foundation for an exciting research program that spans cognitive neuroscience and SLA. Future directions include additional analyses on data already collected as part of this dissertation project, as well as other planned projects that can address the validity of mini-language research as a tool for understanding L2 learning, and more deeply answer fundamental questions in SLA research.
First, this experiment has provided a wealth of data that has the potential to be very informative in many areas of SLA and cognitive neuroscience research. In this dissertation, I have attempted to provide a general overview of analyses that most directly answer my research questions and provide a foundation for further analyses. Additional analyses will include simple analyses that were not of particular relevance here, such as RT analyses for production, a more fine-grained breakdown of sentence production according to language domain, and analyses of the written GJT, think-alouds, free response tasks, and debriefing questionnaires. Regarding fMRI, separate analyses will be run for word conditions (nouns, verbs, adjectives) to probe whether different regions are involved in learning different parts of speech (Mestres-Missé et al., 2010), and the relationship between exposure/proficiency and activation will be analyzed separately for each sentence condition. Most importantly, I will explore more sophisticated analyses to examine changes in neural activation over time and according to proficiency, and attempt to examine the neural bases of language learning with delta analyses.

A very interesting aspect of L2 learning that was tested in this dissertation, but has not yet been analyzed, is language retention. Most subjects returned to complete additional Mini-Basque tasks about a month after training, and several of them completed some of these tasks in the MRI scanner. Language retention is an area of L2 research that is understudied, especially at the neural level, but previous findings suggest that examining retention can tell us a lot about L2 development and processing (Morgan-Short, Finger, et al., 2012), particularly for language abilities that are maintained over time. Retention is a particularly important aspect of SLA research, as the goal of L2 learners is not just to attain high proficiency, but to maintain it.
Further analyses will be conducted on data from the retention sessions for the Mini-Basque learners, both alone and compared to the training sessions (i.e., Days 1 – 3).

This project also involved the collection of data on various cognitive abilities (Day Ø), but the analysis of these data was not within the scope of this dissertation. Individual differences in cognitive abilities are consistently evoked as important factors in determining L2 learning processes and outcomes (Dörnyei, 2006), though little is known about the relationship between individual differences and the neural correlates of L2. Understanding this relationship can help determine how different kinds of learners perform on L2 tasks and process their L2 at a neural level. For example, learners who demonstrate good procedural memory abilities might be more likely to engage implicit learning mechanisms, like the basal ganglia, whereas learners who demonstrate better working memory or declarative memory abilities might be more likely to rely on declarative memory structures when learning grammar. These and other questions related to individual differences in L2 acquisition will be explored in further analyses.

A final data set that was not analyzed in this dissertation comprises the tests and questionnaires performed at the end of training and testing, which are based on similar materials frequently used in SLA. The written GJTs can provide a wealth of information about metalinguistic knowledge and awareness, which can be complemented by results from the debriefing questionnaires. Additionally, the free response task can demonstrate the extent to which learners are able to use the mini-language in a more natural, communicative setting than the very constrained production task. While these tasks are still not completely natural assessments of language knowledge, they are more commonly used in SLA and can thus be directly compared to the literature in this field.
In addition to analyses pertaining to existing data, future directions of this research program also include designing studies to answer other related questions. First, as mentioned above in Section 4.3, the use of a mini-language as a model for a natural language is a limitation of this study, as the extent to which mini-language L2 processing can generalize to natural language processing is unknown. In an attempt to validate the use of Mini-Basque as a model for language, native speakers and L2 speakers of Basque can be tested on the mini-language. By comparing the behavioral and neural results from Mini-Basque learners to L1 Basque speakers, we can directly examine whether mini-language learners exhibit language-related brain patterns, as well as the extent to which these patterns approximate those of native speakers. By comparing the Mini-Basque learners to L2 Basque speakers of varying proficiency levels, we can directly examine whether the two groups exhibit similar brain patterns. If they do, particularly at similar levels of performance, this would help validate the use of a mini-language and demonstrate its generalizability as a model for L2 acquisition, and thus encourage the use of such models in future research. This sort of validation would make it much more feasible to conduct continuous, longitudinal studies of L2 learning and processing, which should continue to fill in gaps regarding the full trajectory of this process.

Another direction of future research pertains to the instructional conditions used to train learners on Mini-Basque. As described in Section 2.2.3.1, the training in this study is consistent with implicit training conditions in SLA. However, a good deal of SLA research is dedicated to understanding the efficacy of different instructional conditions, and the extent to which different conditions lead to implicit or explicit learning and knowledge. While most behavioral evidence suggests that different conditions do differentially affect learning and knowledge (Norris &
Ortega, 2001; Spada & Tomita, 2010), recent work on artificial language learning and ERPs has shown that with enough training, learners trained under implicit and explicit conditions demonstrate similar behavioral performance, but exhibit different neural patterns, both at the end of training to high proficiency (Morgan-Short, Steinhauer, et al., 2012), and after several months with no exposure to the artificial language (Morgan-Short, Finger, et al., 2012). In both cases, learners in the more implicit group exhibited more native-like brain patterns, suggesting an advantage for a more immersive learning condition. These findings underscore the importance of incorporating neural measures to gain a better understanding of implicit and explicit instruction, learning, and knowledge in L2. Future studies of Mini-Basque could employ a variety of training conditions, modified in areas like instruction and feedback, to investigate these variables in Mini-Basque learners.

Taken together, analyses and additional studies that target language retention, individual differences, tasks probing L2 knowledge and awareness, validity and generalizability of the mini-language, and different learning conditions can provide a great deal of information about the neural and behavioral bases of L2 learning and processing. The goals of applied linguistics and cognitive neuroscience are distinct – the former aims to characterize and promote L2 acquisition, and the latter aims to understand the neural underpinnings of L2 learning and processing – but they are not incompatible. Going forward, researchers should not lose sight of the importance of combining behavioral approaches from SLA with neuroimaging techniques from cognitive neuroscience. Neuroimaging techniques may help us understand behavioral performance and how certain proficiency levels can be achieved, but ultimately, the goal of language learners is not to exhibit certain brain patterns, but to achieve high proficiency in the
L2. As these fields are just beginning to merge, it is still not completely clear how our knowledge about the neural bases of language can work to promote actual L2 learning, but further studies and analyses in this line of research, informed by both fields, are likely to bring us closer to solving this problem.

This dissertation project examined the longitudinal and continuous learning of a mini-language over the course of several days of training. While the learners in this study, on average, did not learn the full mini-language system, they still demonstrated reasonable improvements over the course of learning. This study is therefore the first, to the author’s knowledge, to investigate longitudinal and continuous learning of natural language stimuli using behavioral and fMRI techniques, even to moderately high proficiency. The results from this study demonstrate changes in the neural underpinnings of L2 acquisition as a function of exposure and proficiency, which has implications for neurocognitive theories of L2 acquisition. Specifically, the results from this study seem to lend the most support to the DP model of language. However, further analyses are needed to continue to contribute to our understanding of the cognitive neuroscience of adult L2 acquisition and processing. This study also has implications for SLA regarding age effects in L2 learning, the relationship between L1 and L2 processing in the brain, the recruitment of implicit and explicit mechanisms in L2 learning and processing, and the operationalization of proficiency.

This study has set the stage for an exciting research program based on using mini-language models and continuous fMRI testing. The relationship between individual differences in performance was tested in this study, but understanding why certain subjects perform better than others, and how that might influence their neural processes, should prove very insightful.
The aptitude data collected as part of this dissertation project can begin to answer these questions. Additionally, manipulating learning conditions could be very informative regarding the extent to which different learning environments shape neural processing. Furthermore, providing learners with explicit instruction prior to training may help improve performance outcomes, and therefore allow for the continuous investigation of L2 learning to high proficiency, which was not achieved in this study. This research should have important implications for research in second language acquisition, language pedagogy, cognitive psychology, and cognitive neuroscience.
# Appendix A

Sample Word Production Answer Sheet (Day 1 of 4)

<table>
<thead>
<tr>
<th>Order</th>
<th>Condition</th>
<th>Word</th>
<th>Correct?</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>noun</td>
<td>txakurra</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>adjective</td>
<td>horia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>verb</td>
<td>ikusi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>noun</td>
<td>zaldia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>verb</td>
<td>usaindu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>adjective</td>
<td>gorria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>noun</td>
<td>hartz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>noun</td>
<td>behia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>adjective</td>
<td>beltza</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>verb</td>
<td>miazkatu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>verb</td>
<td>bultzatu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>adjective</td>
<td>urdina</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>adjective</td>
<td>gorria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>verb</td>
<td>bultzatu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>noun</td>
<td>txakurra</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>adjective</td>
<td>urdina</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>verb</td>
<td>miazkatu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>adjective</td>
<td>beltza</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>noun</td>
<td>hartz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>noun</td>
<td>zaldia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>adjective</td>
<td>horia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>noun</td>
<td>behia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>verb</td>
<td>usaindu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>verb</td>
<td>ikusi</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix B

Sample Sentence Production Answer Sheet (Day 1 of 4)

<table>
<thead>
<tr>
<th>Run</th>
<th>Order</th>
<th>ID</th>
<th>Subject</th>
<th>Adj (S)</th>
<th>Det (S)</th>
<th>Object</th>
<th>Adj (O)</th>
<th>Det (O)</th>
<th>Verb</th>
<th>Aux</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>667</td>
<td>Txakur</td>
<td>gorri</td>
<td>ak</td>
<td>hartz</td>
<td>urdin</td>
<td>a</td>
<td>bultzatu</td>
<td>du</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>684</td>
<td>Behi</td>
<td>hori</td>
<td>ek</td>
<td>behi</td>
<td>urdin</td>
<td>ak</td>
<td>bultzatu</td>
<td>dituzte</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>672</td>
<td>Behi</td>
<td>gorri</td>
<td>ak</td>
<td>hartz</td>
<td>gorri</td>
<td>a</td>
<td>ikusi</td>
<td>du</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>654</td>
<td>Behi</td>
<td>beltz</td>
<td>ek</td>
<td>hartz</td>
<td>a</td>
<td></td>
<td>usaindu</td>
<td>dute</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>635</td>
<td>Txakur</td>
<td></td>
<td>ak</td>
<td>hartz</td>
<td>ak</td>
<td></td>
<td>ikusi</td>
<td>ditu</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>696</td>
<td>Zaldi</td>
<td>gorri</td>
<td>ek</td>
<td>zaldi</td>
<td>beltz</td>
<td>a</td>
<td>bultzatu</td>
<td>dute</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>695</td>
<td>Zaldi</td>
<td>urdin</td>
<td>ek</td>
<td>hartz</td>
<td>gorri</td>
<td>ak</td>
<td>miazkatu</td>
<td>dituzte</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>645</td>
<td>Behi</td>
<td></td>
<td>ek</td>
<td>txakur</td>
<td>gorri</td>
<td>ak</td>
<td>bultzatu</td>
<td>dituzte</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>610</td>
<td>Zaldi</td>
<td>hori</td>
<td>ak</td>
<td>txakur</td>
<td>a</td>
<td></td>
<td>usaindu</td>
<td>du</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>673</td>
<td>Txakur</td>
<td>beltz</td>
<td>ek</td>
<td>behi</td>
<td>hori</td>
<td>a</td>
<td>ikusi</td>
<td>dute</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>625</td>
<td>Zaldi</td>
<td>beltz</td>
<td>ak</td>
<td>txakur</td>
<td>beltz</td>
<td>ak</td>
<td>usaindu</td>
<td>ditu</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>660</td>
<td>Zaldi</td>
<td>beltz</td>
<td>ak</td>
<td>zaldi</td>
<td>gorri</td>
<td>ak</td>
<td>bultzatu</td>
<td>ditu</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>607</td>
<td>Behi</td>
<td>urdin</td>
<td>ek</td>
<td>txakur</td>
<td>hori</td>
<td>a</td>
<td>bultzatu</td>
<td>dute</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>693</td>
<td>Behi</td>
<td>gorri</td>
<td>ek</td>
<td>behi</td>
<td>a</td>
<td></td>
<td>miazkatu</td>
<td>dute</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>638</td>
<td>Hartz</td>
<td>gorri</td>
<td>ak</td>
<td>hartz</td>
<td>urdin</td>
<td>ak</td>
<td>ikusi</td>
<td>ditu</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>642</td>
<td>Behi</td>
<td>ak</td>
<td>behi</td>
<td>beltz</td>
<td>ak</td>
<td>miazkatu</td>
<td>ditu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>679</td>
<td>Hartz</td>
<td>gorri</td>
<td>ak</td>
<td>zaldi</td>
<td>beltz</td>
<td>a</td>
<td>bultzatu</td>
<td>du</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>661</td>
<td>Behi</td>
<td>hori</td>
<td>ak</td>
<td>hartz</td>
<td>a</td>
<td>ikusi</td>
<td>du</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>19</td>
<td>605</td>
<td>Txakur</td>
<td>gorri</td>
<td>ak</td>
<td>behi</td>
<td>gorri</td>
<td>a</td>
<td>ikusi</td>
<td>du</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>606</td>
<td>Zaldi</td>
<td>urdin</td>
<td>ek</td>
<td>behi</td>
<td>urdin</td>
<td>ak</td>
<td>usaindu</td>
<td>dituzte</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>689</td>
<td>Hartz</td>
<td>urdin</td>
<td>ek</td>
<td>behi</td>
<td>hori</td>
<td>ak</td>
<td>usaindu</td>
<td>dituzte</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>633</td>
<td>Behi</td>
<td>beltz</td>
<td>ek</td>
<td>behi</td>
<td>gorri</td>
<td>a</td>
<td>miazkatu</td>
<td>dute</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>631</td>
<td>Txakur</td>
<td>urdin</td>
<td>ek</td>
<td>zaldi</td>
<td>ak</td>
<td>usaindu</td>
<td>dituzte</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>675</td>
<td>Zaldi</td>
<td>urdin</td>
<td>ak</td>
<td>hartz</td>
<td>urdin</td>
<td>ak</td>
<td>usaindu</td>
<td>ditu</td>
<td></td>
</tr>
</tbody>
</table>

Notes:
Appendix C

Written Grammaticality Judgment Test (Version 1 of 4)

Participant No. __________  Judgment score:__________
Date: __________ Correction score:__________

Mini-Language Written Test (v1)

Instructions: Carefully read the sentences below and decide whether each sentence is ‘Good’ or ‘Bad’ in the mini-language. Circle the appropriate word (Good/Bad) to the right of the sentence.

If the sentence is **bad**, please take the following 2 steps:
1. Make any necessary changes in the sentence to make it a correct sentence, and
2. Write an explanation about why the sentence was incorrect.

Here’s an example of how this would work in English:

1. The boy like ice cream.  
   Good / Bad
   The boy likes ice cream. (*corrected sentence*)
   ________________________________
   The sentence is incorrect because when the subject of the sentence is a third person singular subject, you need to add an “-s” to the end of the verb. (*explanation*)

Please work through the test sequentially and do not go back and change your answers.

I am interested in what you say to yourself as you perform this test. In order to gather this information, I will ask you to TALK ALOUD as you go through the test. What I mean by talk aloud is that I want you to say out loud everything that you say to yourself silently. Just act as if you are alone in the room speaking to yourself. Don't try to explain your thoughts, just say whatever you are thinking to yourself. If you are silent for any length of time I will remind you to keep talking aloud.

You may take as much time as you need to complete this test.

1. Behi horiak  
   txakur gorriak  
   miazkatu du.  
   Good / Bad
   ________________________________

2. Behiak hori  
   txakurak  
   bultzatu ditu.  
   Good / Bad
   ________________________________
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.</td>
<td>Zaldi horiek</td>
<td>hartz gorriak</td>
<td>miazkatu dituzte.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Hartzek</td>
<td>zaldi dituzte urdinak</td>
<td>bultzatu.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Behi urdinak</td>
<td>behi beltza</td>
<td>miazkatu du.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Behi urdinak</td>
<td>txakur urdinak</td>
<td>usaindu du.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Behiak</td>
<td>behia urdin</td>
<td>miazkatu du.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Zaldi horiak</td>
<td>behi urdinak</td>
<td>usaindu ditu.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Zaldi gorriak</td>
<td>hartz ditu gorriak</td>
<td>ikusi.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Hartz horiak</td>
<td>txakur gorria</td>
<td>usaindu du.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Zaldiek</td>
<td>hartz beltza</td>
<td>usaindu dituzte.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Good / Bad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Behi beltzak</td>
<td>behi gorriak</td>
<td>miazkatu ditu.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Hartz urdinak</td>
<td>txakur urdinak</td>
<td>ikusi dituzte.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>Txakur horiak</td>
<td>hartzak</td>
<td>bultzatu ditu.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>Zaldi beltzek</td>
<td>hartza</td>
<td>ikusi dute.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>Txakur beltzek</td>
<td>txakur horiak</td>
<td>ikusi dute.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>Txakur urdinak</td>
<td>hartz gorria</td>
<td>miazkatu du.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hartz beltzek</td>
<td>hartz gorriak</td>
<td>bultzatu dute.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---------------</td>
<td>---------------</td>
<td>----------------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>Behi horiek</td>
<td>behi horiak</td>
<td>bultzatu dituzte.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>Hartz horiek</td>
<td>txakur gorriak</td>
<td>usaindu dituzte.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>Behi dituzte gorriek</td>
<td>behi beltzak</td>
<td>bultzatu.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>Txakurek</td>
<td>hartza gorri</td>
<td>usaindu dute.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>Hartzek</td>
<td>hartz gorriak</td>
<td>ikusi dituzte.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.</td>
<td>Zaldi gorriek</td>
<td>txakur gorriak</td>
<td>usaindu ditu.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>Hartzek urdin</td>
<td>hartzak</td>
<td>miazkatu dituzte.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.</td>
<td>Behi gorriek</td>
<td>txakur urdinak</td>
<td>bultzatu dituzte.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. Hartz horiak</td>
<td>txakur horia</td>
<td>miazkatu du.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. Behi horiak</td>
<td>zaldi gorriak</td>
<td>miazkatu ditu.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. Hartzak</td>
<td>txakurak</td>
<td>bultzatu ditu.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. Txakurak</td>
<td>txakur horia</td>
<td>miazkatu ditu.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34. Behi urdinek</td>
<td>hartzak</td>
<td>usaindu dituzte.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35. Zaldi gorriek</td>
<td>zaldi gorria</td>
<td>usaindu dute.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36. Hartz beltzek</td>
<td>hartzak</td>
<td>bultzatu dute.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37. Txakur horiak</td>
<td>behi horia</td>
<td>miazkatu ditu.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38. Hartzak</td>
<td>hartz urdina</td>
<td>usaindu du.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39.</td>
<td>Txakur dute horiek</td>
<td>hartz gorria</td>
<td>ikusi.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.</td>
<td>Hartz beltzek</td>
<td>txakura</td>
<td>bultzatu dute.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41.</td>
<td>Hartz horiak</td>
<td>zaldi beltza</td>
<td>bultzatu dute.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42.</td>
<td>Hartzak</td>
<td>txakur urdina</td>
<td>miazkatu du.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43.</td>
<td>Txakurak urdin</td>
<td>txakura gorri</td>
<td>bultzatu du.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.</td>
<td>Hartz beltzek</td>
<td>zaldi horia</td>
<td>ikusi du.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45.</td>
<td>Hartz beltzek</td>
<td>txakur urdina</td>
<td>ikusi dute.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46.</td>
<td>Zaldi urdinek</td>
<td>behi dute urdina</td>
<td>miazkatu.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47.</td>
<td>Txakur horiek</td>
<td>zaldi beltza</td>
<td>miazkatu dute.</td>
<td>Good / Bad</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D

Debriefing Questionnaire (End of Training)

Subject No.:_________________________ Date: ___________

Please carefully read and answer the following questions regarding the language study you have just completed. Answer the questions in as much detail as possible. You may take as much time as you need to complete this questionnaire.

Please answer the questions **IN ORDER** and **DO NOT GO BACK**.

1. a. When performing the tasks in this study, did you notice any rules or regularities in the mini-language?

   YES       NO

If yes, mark below when you first noticed the rules:

   ____ During *picture-matching* before the 1st acceptability judgment test
   ____ During *picture-matching* after the 1st acceptability judgment test
   ____ During *production*
   ____ During the *written test*
   ____ At another time (please indicate) __________________________________________

b. Were you looking for rules or regularities in the mini-language?

   YES       NO

c. If you can, please explain what sort of rules or regularities you noticed? If needed, give an example to illustrate.

_______________________________________________________________________
_______________________________________________________________________
2. a. Did you notice anything about **word order** in the mini-language?

    YES  NO

If yes, mark below when you first noticed the rules:

- During *picture-matching* before the 1st acceptability judgment test
- During *picture-matching* after the 1st acceptability judgment test
- During *production*
- During the *written test*
- At another time (please indicate) ____________________________________________

b. Were you looking for rules about **word order** in the mini-language?

    YES  NO

c. If you can, please explain what kinds of rules you noticed about **word order**? If needed, give an example to illustrate.

_______________________________________________________________________

_______________________________________________________________________

_______________________________________________________________________

3. a. Did you notice anything about **endings** on some of the words in the mini-language?

    YES  NO

If yes, mark below when you first noticed the rules:

- During *picture-matching* before the 1st acceptability judgment test
- During *picture-matching* after the 1st acceptability judgment test
- During *production*
- During the *written test*
- At another time (please indicate) ____________________________________________

b. Were you looking for rules related to these endings (\(-a, -ak, -ek\)) during the experiment?

    YES  NO
c. If you can, please write some sort of ‘rule(s)’ that would explain how to use these particular endings (-a, -ak, -ek). If needed, give an example to illustrate.

_______________________________________________________________________

_______________________________________________________________________

_______________________________________________________________________

PLEASE GO TO THE NEXT PAGE

4. a. Did you notice anything about the relationship between the endings (-a, -ak, -ek) and other words in the mini-language?

YES

NO

If yes, mark below when you first noticed the rules:

_____ During picture-matching before the 1st acceptability judgment test
_____ During picture-matching after the 1st acceptability judgment test
_____ During production
_____ During the written test
_____ At another time (please indicate) __________________________________________________________________________

b. If yes, which words do you think these endings (-a, -ak, -ek) were related to?

_______________________________________________________________________

PLEASE GO TO THE NEXT PAGE
5. a. Did you notice anything about the relationship between the endings (-a, -ak, -ek) and the following words: *du, dute, ditu, dituzte*?

   YES                      NO

   If yes, mark below when you first noticed the rules:
   _____ During *picture-matching* before the 1st acceptability judgment test
   _____ During *picture-matching* after the 1st acceptability judgment test
   _____ During *production*
   _____ During the *written test*
   _____ At another time (please indicate) ______________________________

b. Were you *looking* for rules related to how these endings (-a, -ak, -ek) and these words (*du, dute, ditu, dituzte*) worked together?

   YES                      NO

c. If you can, please write some sort of ‘rule(s)’ that would explain how these particular mini-language words (*du, dute, ditu, dituzte*) and endings (-a, -ak, -ek) are related. If needed, give an example to illustrate.

   ___________________________________________________________________
   ___________________________________________________________________
   ___________________________________________________________________

6. What kind of information, if any, did these particular mini-language words (*du, dute, ditu, dituzte*) and endings (-a, -ak, -ek) tell you about the rest of the sentence?

   ___________________________________________________________________
   ___________________________________________________________________
   ___________________________________________________________________

7. Is there anything else you would like to tell us or think we should know about the study?

   ___________________________________________________________________
   ___________________________________________________________________
   ___________________________________________________________________
   ___________________________________________________________________
Appendix E

Wrap-Up Questionnaire (End of Training)

Mini-language learning: Wrap up

Subject: _________________________  Date: __________________

Please take a moment to answer some wrap-up questions about this mini-language learning study.

1. Are you CURRENTLY enrolled in a language course?

   Please circle your answer.

   Yes  No

   If you circled yes, please briefly describe your language course in the space provided. For example, what language? What level? How often does the class meet? Any notable features of your instructor’s teaching style?

   ______________________________________________________________
   ______________________________________________________________
   ______________________________________________________________
   ______________________________________________________________
   ______________________________________________________________

2. During several sessions of this study, you have been learning and practicing a mini-language, which is based on a real natural language. Have you tried to find out any more information about the language outside of the study?

   Please circle your answer.

   Yes  No

   If Yes, please continue to question 3. If No, skip to question 4.
3a. Please tell us a little more about when you looked for more information (between which sessions), what information you were interested in knowing, and what you were able to find out.

______________________________________________________________
______________________________________________________________
______________________________________________________________
______________________________________________________________

3b. Do you think that the information you found changed your performance on any of the language training and testing tasks you performed in this study, in and/or out of the MRI scanner? Why or why not?

______________________________________________________________
______________________________________________________________
______________________________________________________________
______________________________________________________________

4. If you know the name of the real, natural language from which the mini-language you learned is derived, please write it below:

______________________________________

Thank you! Please let Kaitlyn know you’ve finished.
Appendix F

Individual Subject Performance on Each fMRI Task

Figure A1. Accuracy in the word-level picture-matching task for each subject.

Figure A2. Accuracy in the sentence-level picture-matching task for each subject.

Figure A3. $D'$ scores for the morpho-syntactic trials in the GJT for each subject.

Figure A4. $D'$ scores for the syntactic trials in the GJT for each subject.
References


Hernandez, A. E., & Meschyan, G. (2006). Executive function is necessary to enhance lexical processing in a less proficient L2: Evidence from fMRI during picture naming. *Bilingualism, 9*(02), 177. doi:10.1017/S1366728906002525


208
doi:10.1017/S0272263105050084


Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.


doi:10.1016/S1053-8119(03)00170-8


doi:10.1523/JNEUROSCI.2220-04.2004


doi:10.1111/j.1467-9922.2006.00361.x


234


Yokoyama, S., Kim, J., Uchida, S., Miyamoto, T., Yoshimoto, K., Riera, J., … Kawashima, R. (2009). Left middle temporal deactivation caused by insufficient second language word
