NEURAL CORRELATES OF LEXICAL INTERACTION IN ADULT SECOND LANGUAGE LEARNERS

A Thesis in

Psychology

by

Angela M. Grant

© 2013 Angela M. Grant

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2013
The thesis of Angela M. Grant was reviewed and approved* by the following:

Ping Li  
Professor of Psychology and Linguistics  
Thesis Adviser

Janet Van Hell  
Professor of Psychology and Linguistics

Rick Gilmore  
Associate Professor of Psychology

Suzy Scherf  
Assistant Professor of Psychology

Melvin Mark  
Professor of Psychology  
Head of the Department of Psychology

*Signatures are on file in the Graduate School.
ABSTRACT

This study uses functional magnetic resonance imaging (fMRI) to address two main issues in second language (L2) research: how semantic access in the L2 interacts with L1 semantic access, and how these processes may be influenced by individual differences in non-linguistic domains. Models of bilingual processing such as the inhibitory control (IC) model and the revised hierarchical model (RHM) suggest that (a) bilinguals must consistently use IC in comprehension and production, and (b) highly proficient learners access concepts directly while less proficient learners access concepts through the L1, respectively. The neural implication of these models is that less proficient bilinguals, compared with highly proficient bilinguals, will require more effort to inhibit their L1 in order to successfully retrieve words in the L2.

Based on these models, we expected that lower proficiency learners would more strongly activate IC areas than areas associated with semantic retrieval when processing their L2 compared with their L1. Higher proficiency learners, by contrast, would utilize networks focused on semantic/conceptual retrieval rather than inhibitory control. Additionally, we expected individual differences in the non-linguistic domains of working memory (WM) and IC to moderate these proficiency-based differences, such that high WM/high IC participants would use show less activation in control areas than their proficiency-matched peers. To examine this, participants complete measures of proficiency, WM, and IC that were regressed on BOLD responses to a lexical decision task. Our results largely supported our predictions, and are discussed in light of the IC and RHM models, as well as brain-based models of L2 processing.
# TABLE OF CONTENTS

LIST OF TABLES ......................................................................................................................... v

LIST OF FIGURES ........................................................................................................................ vi

Chapter 1: Neural Correlates of Lexical Interaction in Adult Second Language Learners .......... 1

Theories of Second Language Development ............................................................................... 2
Neural Markers of Lexical Interaction between the L1 and L2 .................................................. 5
Individual Differences in Language Learning ........................................................................... 12
Goals of the Current Study ......................................................................................................... 14

Chapter 2: Methods ...................................................................................................................... 17

Participants ................................................................................................................................. 17
Materials .................................................................................................................................. 18
Procedure ................................................................................................................................. 19
Data Analysis ............................................................................................................................ 21

Chapter 3: Results ...................................................................................................................... 25

Behavioral Results ..................................................................................................................... 25
fMRI Results .............................................................................................................................. 27

Chapter 4: General Discussion .................................................................................................. 40

Semantic Access in the L1 and the L2 ....................................................................................... 42
Influence of Individual Differences on L2 Semantic Access ..................................................... 44

References .................................................................................................................................. 49
LIST OF TABLES

Table 1: *Stimulus words for each experimental condition in the lexical decision task.* ....................19

Table 2: *Session 1 Measures* ..............................................................................................................25

Table 3: *ROI Analyses at Session 2 and Session 3* ...........................................................................29
LIST OF FIGURES

Figure 1. Rodríguez-Fornells et al.’s (2009) model. .................................................................10

Figure 2. Language switching interaction between the IFG and MTG as reported by Nakamura et al. 2010. ..............................................................................................................................12

Figure 3. Proficiency as measured by self-report and performance on the Spanish version of the Peabody Picture Vocabulary Test (Test de Vocabulario en Imagenes Peabody)................18

Figure 4. Time course of a trial in the lexical decision task........................................................21

Figure 5. Session 2 and 3 Group Maps ..........................................................................................33

Figure 6. Control Network at Sessions 2 and 3 ..........................................................................35

Figure 7. Control and meaning network interface at Sessions 2 and 3.....................................38
Chapter 1: Neural Correlates of Lexical Interaction in Adult Second Language Learners

More than half the world’s population is bilingual (Grosjean & Li, 2012; Chapter 1). This statistic is so large partly because the definition of a “bilingual” ranges widely: from the prototypical idea of someone raised from birth speaking two languages, to the struggling undergraduate fulfilling a language requirement, to the businessman learning the language of the country to which he has been transferred (Bialystok & Hakuta, 1994). In the United States, however, most bilinguals fall somewhere around the second category: classroom learners (Grosjean, 1982). Consequently, research that focuses on this most prevalent type of second language (L2) learner, the classroom learner, has the benefit of being able to provide ecologically valid observations about how learning occurs in these individuals.

For any learner, however, one of the first basic steps in acquiring a language is to learn the vocabulary. Not surprisingly, much of the psycholinguistic literature on bilingualism and L2 acquisition has been devoted to understanding how words in one’s second language are learned and processed, and how those processes compare to processing in one’s first language (Potter, So, Von Eckardt, & Feldman, 1984; Kroll & Stewart, 1994). With the advent of neuroimaging technology, a new stream of research began to ask questions concerning word learning in the L2. Would L2 processing occur in the same brain areas as the L1? Would age of acquisition (AoA) or proficiency modulate the types of areas activated (Chee, Tan, & Thiel, 1999)? How would lexical processing compare with syntactic processing (Ullman, 2001; Clahsen & Felser, 2006)? Yet, few of these studies, behavioral or neuroimaging, attempted to longitudinally track the development of the processing networks in question (but, see McLaughlin, Osterhout & Kim,
2004; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2007) and consequently much of our knowledge is based on cross-sectional research that does not allow us to observe the subtleties and individual differences in the development of L2 processing.

The current research will address concerns of both ecological and developmental validity while providing further insight into the process of L2 learning. The current study expands on previous research by using neuroimaging methodology, specifically functional magnetic resonance imaging (fMRI), to achieve two main goals: to track the developmental changes associated with semantic access in the L1 and L2, and investigate how these linguistic processes may be influenced by individual differences in the non-linguistic domains of working memory and cognitive control.

**Theories of Second Language Development**

One of the best-known models of bilingual language development, the Revised Hierarchical Model, hereafter referred to as the RHM (Kroll & Stewart, 1994), is based on the work of Potter et al. (1984). Potter et al. proposed two possible routes through which L2 learners access L2 words: word association and concept mediation. The word association route presupposed that learners associated the two word forms directly and consequently accessed the L2 word meaning by first accessing the meaning of its translation equivalent in the L1. In contrast, the concept mediation hypothesis suggested that learners would associate L2 words directly with concepts, rather than having to detour through the L1. Subsequent experiments (e.g. Kroll & Curley, 1988) provided evidence for both hypotheses, and based on the pattern of these results Kroll and Stewart (1994) created the Revised Hierarchical Model (RHM). Kroll and Stewart’s model added on to Potter et al.’s (1984) original hypotheses by proposing a developmental timeline. Indeed, the RHM predicts that learners will first associate the two word
forms, and then gradually develop and strengthen separate links to the conceptual store, which is shared between both languages.

The RHM, however, is a model based on behavior. Given the neurocognitive nature of the current study, a description of the primary neurological theories of L2 development is also pertinent. There are two basic hypotheses concerning the neural basis of L2 learning (Abutalebi & Green, 2007). The differential representation hypothesis suggests that AoA is the most important factor and that second languages are represented and processed differently than the L1. This hypothesis has been espoused by researchers such as Ullman (2001, 2004), who suggest that because L2 grammar is learned explicitly, L2 grammar is subserved by declarative memory, situated in the medial temporal regions, the same area as L1 and L2 vocabulary. This mode of syntactic processing contrasts with the proposed L1 pattern, which Ullman posits to be subserved by procedural memory which is instantiated in fronto-basal ganglia circuits. Consequently, L2 production would be mediated by explicit, metalinguistic processing as opposed to the implicit way in which one’s L1 is produced. Clahsen and Felser (2006) put forward a similar proposal, called the “shallow structure hypothesis”, that claimed that the syntactic representations of adult L2 learners are shallower and less detailed than those of native speakers, although their computation of lexical-semantic cues are essentially the same. These conclusions were based on research that had found that highly proficient L2 learners showed ERP responses similar to native speakers during morphological processing, their processing of syntactic dependencies appeared to rely more on lexical-semantic and pragmatic information, rather than syntactic gaps.

Green (2003) proposed an alternative, the convergence hypothesis. His hypothesis suggests that with increasing proficiency L2 learners represent and process the L2 similarly to native speakers. The convergence hypothesis is based on the idea that the computational
requirements to produce and store words and to put together sentences are different, and consequently the most efficient solution to L2 production is to use the same pre-specified circuits as those involved in the L1. Abutalebi and Green (2007) argue in their review paper that recent evidence appears to support the convergence hypothesis, as a number of studies (Kim et al., 1997; Chee et al. 1999; Hernandez et al., 2001) have found activation of identical areas in both the L1 and L2. Additionally, Abutalebi (2008) argues in favor of convergence, although he notes that lower proficiency learners often utilize prefrontal, caudate, and anterior cingulate cortex (ACC) resources in addition to traditional L1 resources.

Recent neuroimaging and computational research (Hernandez & Li, 2007), however, suggests that neither position is entirely correct. Instead, Hernandez and Li suggest that for vocabulary acquisition, specifically, proficiency may be more important than AoA, whereas for grammar acquisition, AoA may play a more important role than proficiency. This claim is based partly on the results of Wartenburger et al. (2003). Wartenburger et al.’s study included three participant groups: Early Acquisition High Proficient (EAHP), Late Acquisition High Proficient (LAHP) and Late Acquisition Low Proficient (LALP). All participants were scanned while monitoring sentences for syntactic or semantic violations. Wartenburger et al.’s results showed that semantic processing occurred differentially for high and low proficiency groups (e.g. activated different networks), whereas for syntactic processing it was the early and late acquisition groups whose processing networks differed. Additionally, Hernandez and Li’s (2007) idea that proficiency is more important for lexico-semantic processing while AoA is more relevant for syntactic processing aligns well with Ullman’s (2001, 2004) declarative-procedural model, described above. The logic of the current study integrates both the neurological theories of Green (2003), and Hernandez & Li (2007) with the behaviorally based developmental
approach of the RHM (Kroll & Stewart, 1994), as will be explained further in the “Goals of the Current Study” section.

**Neural Markers of Lexical Interaction between the L1 and L2**

The advent of neuroimaging research has offered a new outlook for examining lexical interaction in second language learners. Researchers utilizing neuroimaging methods are able to combine traditional behavioral measures, such as reaction time or accuracy, with the neuroimaging data to arrive at conclusions based on converging evidence. In the case of fMRI, we are able to combine behavioral measures with data on the blood oxygen level dependent (BOLD) signal, which tracks the hemodynamic response—an indicator of activity. In the discussion of lexical interaction, many fMRI researchers have taken the approach of examining different areas associated with the two languages, although one may also examine other characteristics of the BOLD signal, as will be discussed shortly.

**Lexico-semantic access of the L1 and L2.** Although the processes involved in selecting one of a bilingual’s two languages are intricately involved with those that allow a single word to be activated, I will separate the two issues here for the sake of clarity. Beginning with the latter issue of lexico-semantic access, Dijkstra and Van Heuven (2002) have proposed a pertinent model of bilingual word comprehension, the BIA+ model. The BIA+ model is divided into two parts, the Word Identification System (WIS) and the Task/Decision (TD) system. Within the WIS, upon seeing a word, bilinguals activate sublexical orthography and phonology, which can influence each other. Activation from these nodes feeds forward (and potentially back) to the lexical orthography and phonology nodes, which are also interconnected. The activation from the lexical orthography and phonology systems feed forward to the language nodes (which do not feed back) and to the combined semantic system (which can feed back). The sum of
information from the WIS is then fed forward to the TD system, which receives continuous input from the WIS but also processes the influences of the non-linguistic context. It is the TD system, not the WIS that ultimately allows the participant to make the final decision required of them (e.g. language membership in a language specific lexical decision task).

One of the early studies to examine lexico-semantic access using fMRI showed that BOLD activation in response to a word generation task (i.e. generate a word using the following stem: *cou*) did not differ in location between the first (Mandarin Chinese) and second (English) languages (Chee et al., 1999). Furthermore, they did not find differences between early (exposed to the L2 before age 6) and late bilinguals (exposed to the L2 after age 12). The authors concluded that their results suggest a shared substrate for first and second language processing, but left open the possibility that the processing itself may differ between the L1 and L2. Additionally, Chee et al. noted that differential organization of the L1 and L2 may still be possible, given that the peak activation in response to the word generation task was in prefrontal regions (BA 44/45 and 46/9) associated with “higher, lexical, and generative aspects of language” (3055). Although their description is vague, the nature of the task could involve inhibiting different possible choices that would complete the stem, which would explain the preponderance of prefrontal activation.

More recent studies, however, have taken more subtle approaches to evaluating differences in lexico-semantic access between the two languages. Rather than focusing on if the two languages used contrasting areas, as Chee et al. (1999) did, Yokoyama et al. (2009) compared relative activation in the left middle temporal cortex. In their study, Yokoyama et al. asked Chinese learners of Japanese, as well as native Japanese speakers, to do a lexical decision task. Like Chee et al. (1999), their participants activated the left inferior frontal gyrus, although
Yokoyama et al.’s (2009) L2 learners showed higher activation than the native Japanese learners. Additionally, however, Yokoyama et al. noted that activation in the left middle temporal gyrus (LMTG) was decreased in L2 learners compared with native speakers. This is particularly interesting because LMTG activation is typically only found in response to words, but not pseudo-words, suggesting that it functions in accessing meaning (Van Heuven et al., 2008). Yokoyama et al. (2009) concluded from this that L2 learners have difficulty in accessing meaning in the L2 because lexical information may not yet be adequately encoded in memory.

Unfortunately, there are a number of potential problems with Yokoyama et al.’s study. First, although they compared the activation of the learners with that of native speakers, they did not include a condition in the first language of the learners (Chinese), which fails to account for within-participant variability between processing in their first and second languages. This is especially worrying considering their small (10 participants per group) sample. Furthermore, although the experimenters used three different word types (nouns, simple active and simple passive verbs) in addition to the pseudo-words, the nature of their design did not allow for an analysis of the responses to each word type. Passive constructions are commonly considered more difficult than active word constructions (Marchman, Bates, Burkhardt, and Good, 1991), thus it is possible that word type also influenced the results. Unfortunately, Yokoyama et al. (2009) did not give a post scan questionnaire, and so it is impossible answer to the questions raised by these problems. The current study addresses the questions raised by Yokoyama et al.’s study by adjusting several aspects of the study design (specified in the Goals of the Current Study section).

**Executive control of linguistic interaction.** The following studies take two different approaches in examining how bilinguals select their languages. One examines neural markers of
lexical interaction by using language-ambivalent stimulus materials, while the others use more traditional materials but focus their predictions and analyses on the connectivity between different areas involved in the process. In the first approach, studying how interlingual homographs (words that are shared between languages) are processed can be more informative than studying language specific words alone. Interlingual homographs are words that are orthographically similar between two languages but differ in their meaning. An example from Spanish and English is the word pie, which means foot in Spanish and pie in English. Because homographs by nature cause activation of the two languages, comparing activation in response to homographs with activation in response to unambiguous L2 words allows the current study to evaluate the idea that early learners use the L1 to access meaning in their L2.

In fact, homographs provide a unique opportunity to examine conflict caused by competing activation of the two languages, and have been used in the past to support the hypothesis that bilinguals control their two languages through inhibitory processes (Macizo, Bajo, & Martín, 2010; Van Heuven, Schriefers, Dijkstra, & Hagoort, 2008). Van Heuven et al. (2008) recently extended this line of research by using fMRI methodology. Their study, which was composed of two experiments, asked bilinguals to either perform a generalized lexical decision task or a language specific lexical decision task. Both experiments used interlingual homographs, and consequently one would expect to observe stimulus-based conflict in both, but only in the language specific lexical decision task would one expect to observe response-based conflict, or conflict due to the need to suppress the other language in order to provide an answer. Thus, they were able to isolate areas related specifically to both stimulus-based conflict, and response-based conflict. The left inferior prefrontal cortex, or LIPFC, responded to stimulus-based conflict while the anterior cingulate cortex (ACC) and the caudate were active in response-
based conflict. Knowing which areas tend to activate, however, is only the first step in understanding the neurological instantiation of language skills. The next step is to examine how these areas interact.

In addition to examining how activation in different areas can correlate with performance on behavioral tasks, newer connectivity-based methodology allows researchers to analyze the interactions between these areas, both through non-directional correlations (functional networks) and through directed connections (effective networks). Additionally, some effective techniques allow for analysis of lagged relationships—the effect of Area X at Time 1 on Area Y at Time 2—as well as contemporaneous relationships—the effect of X on Y at Time 1.

A unique example of recent work in this area is the functional model of Rodríguez-Fornells, Cunillera, Mestres-Missé, and de Diego-Balaguer (2009). (An illustration of Rodríguez-Fornells et al.’s model can be found in Figure 1.) Rodríguez-Fornells et al.’s model proposes that language learning consists of interaction between three main streams: a dorsal audio-motor interface, a ventral meaning inference interface, and an episodic-lexical interface.

For the purposes of the proposed study, a description of the ventral meaning and episodic-lexical interfaces are most pertinent. The ventral meaning inference interface consists of two main divisions, one dedicated to conceptual storage and retrieval and one dedicated to controlling semantic retrieval. Rodríguez-Fornells et al. presumed that the ventral inferior frontal gyrus (vIFG) moderated the latter division. This is the same area noted by Chee et al. (1999) and Yokoyama et al. (2009) to have been especially active in L2 learners. By contrast, the conceptual storage and retrieval division assumes the interaction of a number of areas, including the anterior and posterior ITG (inferior temporal gyrus), the MTG (medial temporal gyrus), and the aTP (anterior temporal pole). Although Rodriguez-Fornells et al. do not specifically outline the roles
of the ITG and aTP, they indicate that these areas are part of the ventral language stream proposed by Hickock and Poeppel (2007). Additionally, they define the role of the MTG as a supramodal semantic processing area that is involved in storing long-term conceptual knowledge, lexical-semantic processing and semantic integration. Finally, the MTL (medial temporal lobe) is thought to be responsible for storing new words and their contextual correlates in what Rodríguez-Fornells et al. (2009) called the episodic-lexical interface. The current study seeks to elucidate the nature of these proposed connections, especially between the vIFG and MTG.

Results from a dynamic causal modeling analysis by Nakamura et al. (2010) elaborate on the predictions of Rodríguez-Fornells et al. (2009). Dynamic causal modeling (DCM), a type of effective connectivity analysis proposed by Friston, Harrison, and Penny (2003) attempts to model the neuronal response. Nakamura et al. (2010) used these analyses on the results of a mixed-language masked priming task to examine language dominance in bilinguals. The
analyses found that the relationship between the IFG and the MTG is bidirectional and that
greater positive coupling between the IFG and MTG occurs for L2 than L1 trials for both
forward and backward connections. Furthermore, this effect interacts with trial type, suggesting
inhibition of the L1 and excitation of the L2 for non-switch trials, but inhibition of both
languages for switch trials, as seen in Figure 2. However, DCM is limited in that it does not
capture lagged relationships between ROIs well (Smith et al., 2011). The proposed study will
attempt to replicate these results using an extended unified Structural Equation Modeling
(euSEM) approach. This approach was developed by Gates, Molenaar, Hillary, and Slobounov
(2011) and builds upon the unified SEM approach put forward by Kim et al. (2007), and allows
for more accurate estimation of the lagged as well as contemporaneous relationships between
ROIs. Additionally, while DCM can measure both contemporaneous and time-lagged effects,
these effects must be built into a confirmatory model. In contrast, euSEM allows for exploratory
analyses and creates the best fitting model from all possible models. Consequently, the euSEM
approach in the proposed study will provide novel information about the functioning of these
language networks.
Individual Differences in Language Learning

The second goal of the proposed study, in addition to examining how L2 processes of semantic access interact with the L1, is to investigate how individual differences in the non-linguistic domains of working memory and cognitive control influence L2 semantic processing. Working memory, according to Baddeley (2012) combines both active storage and manipulation of information. Each of these abilities is then further assigned, in Baddeley’s model, to different components. Manipulation is within the purview of the central executive (hence the term executive control, otherwise known as cognitive control), while the phonological loop, episodic buffer, and visuo-spatial sketchpad are all involved in the active storage of different modes of information. As observed in the previous section, executive control is clearly involved in the process of selecting a language in proficient bilinguals (Van Heuven et al, 2008; Nakamura et al, 2010; Rodriguez-Fornells et al 2009). In addition, Bialystok (2009) has shown that bilingualism is correlated with better executive control. She further suggests that a lifetime of continually needing to inhibit another language is the source of this enhancement, although others (e.g. Hilchey and Klein, 2011) suggest that the bilingual advantage in executive control is more
generalized and less centered around inhibitory control ability. The current research builds from these findings by examining the opposite direction of the correlation: the impact of domain general abilities in working memory as well as inhibitory control on second language processing and acquisition.

Kroll and Linck (2005) provide a review of this relationship and suggest that successful attainment of the L2 relies on cognitive skills in addition to the development of lexical and syntactic L2 representations. They observed that participants who scored high on a memory span task were faster translators and processed L2 words more conceptually than their proficiency matched peers. Specifically, Tokowicz, Michael, and Kroll (2004), observed an interaction between study-abroad experience and working memory such that participants with more study abroad experience and higher working memory were more likely to make meaningful rather than unrelated errors during a translation task. A more recent study, conducted by Linck, Hoshino, and Kroll (2008) found that enhanced working memory and executive control were correlated with reduced cross-language activation—as measured by cognate facilitation—during naming.

Additionally, a recent study by Yang, Swick, Watkins, and Li (2012) found a correlation between working memory ability and success in learning non-native lexical tones. Furthermore, they noted that success in the behavioral tone discrimination task was associated with a network of activation including the bilateral middle frontal gyrus, the middle superior temporal gyrus, and the cerebellum. Whether these associations are specific to the type of language task attempted (i.e. a phonological task in the Yang et al. study compared to the visual task in the present study) has yet to be determined. The current study will promote this growing area of research by examining the relationship between working memory and executive control (as measured by a
letter-number sequencing task and the flanker task, respectively) and the BOLD response elicited by homographs and unambiguous L2 words.

Goals of the Current Study

In sum, the current study sought to accomplish two goals: to track the developmental changes associated with semantic access in the L1 and L2, and investigate how they may be influenced by individual differences in non-linguistic domains. In order to fulfill the first goal and allow for comparisons of the relationship between the L1 and L2, the current study analyzed the BOLD response patterns to interlingual homographs, as well as unambiguous L1 and L2 words. By asking participants to categorize both interlingual homographs as well as unambiguous L1 and L2 words, the current study addresses its goal of comparing how lexical processing and semantic access occurs in the first and second languages. Additionally, single-language lexical decision tasks have previously been shown to elicit conflict and consequently activation of language control areas. By regressing the influences of individual differences in working memory and executive control on these data to determine their respective contributions, this study meets its second goal.

In addition to the aforementioned goals, the current study addressed issues ignored by previous research. For example, comparing BOLD data from learners at two time points (once in the fall and once in the spring) over one academic year, will insure both ecological and developmental validity of the results. Furthermore, using an event-related design allows for the analysis of each of the different stimulus types independently. The current study also included an L1 condition for use as a within-subjects comparison and a post scan questionnaire to evaluate pre-existing knowledge of the stimuli. Additionally, the current study used connectivity analyses
to inform our knowledge not only about the location and relative activation of the regions of interest (ROIs) in L2 learning, but also about the nature of the relationships between them.

Based on the analyses and hypotheses of Hernandez and Li (2007), one would expect to be able to observe development towards convergence in a typical undergraduate sample of adult second language learners. The following hypotheses are based on that pattern of results:

1) I expected that lower proficiency learners would engage inhibitory control areas, including the middle frontal gyrus (MFG), medial frontal gyrus (MeFG), inferior frontal gyrus (IFG), anterior cingulate cortex (ACC) and caudate (CAD), in addition to areas associated with semantic retrieval, such as the LMTG, for both homographs and unambiguous L2 words. Furthermore, I expected to observe lesser activation of the LMTG in low proficient learners based on the results of Yokoyama et al (2009).

2) I expected that higher proficiency learners would utilize networks that center on the LMTG and consequently have less significant activation in the aforementioned control areas for unambiguos L2 words. Furthermore, the L2 networks should resemble the L1 networks in these participants. Interlingual homographs, however, should continue to elicit activation in control areas.

3) I expected that these proficiency-based differences may also be moderated by individual differences in working memory (WM) and inhibitory control (IC) abilities, such that high WM/high IC participants would use networks that rely less on activation in the control areas listed above than their proficiency-matched peers.

The alternative hypothesis was that the participants would show a primarily divergent pattern and would not converge to the pattern of their L1. For example, a participant who made
gains in proficiency but continued to experience increased levels of activation in control areas to
the unambiguous Spanish words compared to the English words would support the divergence
hypothesis. If that were the case, that result would also make a valuable theoretical contribution
to the field. Such a result would suggest that increasing proficiency—the factor posited to
influence the development towards convergence of L1 and L2 processing networks—may not be
the key factor involved in this process. In this case, it is possible that a non-linguistic factor, such
as executive control, could be more important. Bialystok (2009) has suggested that bilingualism
results in the development of higher executive control abilities, but it is also possible that only
students who are skilled inhibitors can become highly proficient bilinguals. The typical
immersion of classroom learners in a primarily L1 environment would only intensify the need for
inhibition, and consequently participants could show a continual pattern of strong LIPFC and
LMTG activation, regardless of proficiency. In this case, I expected to observe similar responses
to both unambiguous L2 words and homographs at each scanner session.
Chapter 2: Methods

Participants

Participants were 36 adult undergraduate native English speakers, although only 24 participated in all three sessions of the experiment. Six participants attrited after the first session, and an additional six attrited after the second session\(^1\).

All participants were right-handed and naïve to the purpose of the experiment. Participants were recruited from 200 to 400 level Spanish courses at the Pennsylvania State University. This range of classes was chosen so that participants would range in proficiency from low-intermediate to high-proficient. Interestingly, this manipulation did not have the intended effect, as there was not a great spread in proficiency scores between participants (see Figure 3). Anecdotally, many of the participants were enrolled in multiple classes at the 200, 300, and 400 levels. This may explain the lack of variation in self-rated proficiency recorded in this experiment.

\(^1\) Of the participants who did not return for the second session, two were dropped for not meeting the inclusion requirements (one was not a native English speaker, and the other was not MRI-safe), and one participant felt claustrophobic after the experience in the mock scanner. The three others did not give reasons for their attrition. Of the additional six that did not return for the 3\(^{rd}\) session, one was studying abroad, and the parents of another were concerned about the safety of the MRI procedure. The remaining four did not give reasons for their choice to drop from the study.
Figure 3. Proficiency as measured by self-report and performance on the Spanish version of the Peabody Picture Vocabulary Test (Test de Vocabulario en Imágenes Peabody).

Materials

Lexical decision task. Stimuli included Spanish and English words, as well as interlingual homographs. All stimuli were equated for length and frequency. Homograph and Spanish words were evaluated using frequency per million statistics from the BuscaPalabras (BPal; Davis & Perea, 2005) database, while English word frequency was evaluated using the Subtlex database (Brysbaert & New, 2009). Spanish words were co-referenced with the textbook from the 200-level classes to insure that the participants were familiar with the stimuli. Example stimuli are found in Table 1. Furthermore, words that were rated by at least 50% of the participants as unfamiliar were later removed from analysis. This procedure resulted in the removal of six words: cutis, mecedora, fuga, pulir, até, and hada.

Semantic judgment task. Prior to completing the lexical decision task, participants completed a short semantic judgment task in the mock scanner to train them on MRI protocol
and to act as an additional measure of proficiency. Specifically, participants were asked to judge if Spanish nouns were living or non-living. All stimuli for this task were in Spanish, and all words were taken from the same textbook as the lexical decision task. None of the stimuli for this task overlapped with the stimuli for the lexical decision task.

Table 1: Stimulus words for each experimental condition in the lexical decision task.

<table>
<thead>
<tr>
<th>Homograph</th>
<th>pie, actual, delfin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unambiguous Spanish</td>
<td>mesa, culpa, diente</td>
</tr>
<tr>
<td>Unambiguous English</td>
<td>rock, calendar, glass</td>
</tr>
</tbody>
</table>

Procedure

Pretesting. At Session 1 all participants completed a language history questionnaire (LHQ; Li, Sepanski, & Zhao, 2006) and handedness questionnaire as well as a battery of computerized tests on Spanish vocabulary (TVIP; Dunn, Lugo, Padilla, and Dunn, 1986), phonological working memory (Letter-Number Sequencing; adapted from Weschler, 1997), and cognitive control (Flanker; adapted from Emmorey, Luk, Pyers, and Bialystok, 2008) before they began the fMRI portion of the study. Participants were screened via the handedness questionnaire and the TVIP for proficiency before returning to participate in the fMRI study.

fMRI procedure. For both Session 2 (November, 2012) and Session 3 (March, 2013) participants were greeted in a separate room where they gave consent to participate in the fMRI portion of the experiment and completed the SLEIC pre-scanning safety questionnaire. Participants then were trained to use the response grips and minimize head movement using the
SLEIC mock scanner. While the experimenter watched and gave commentary on any movement, participants completed a Spanish semantic judgment task. If the participant felt comfortable, he or she would continue to complete the lexical decision task in the real MRI scanner. In the lexical decision task participants were asked to judge whether the words presented are Spanish words. Participants were instructed to respond to homographs (described as “words that could be used in Spanish or English”) as Spanish words. They responded via button-press: 1 for a Spanish word, 2 for a non-Spanish word.

**Design.** Order of language presentation and word type were mixed in a rapid event related design. The study included 3 runs. Each study run included 90 trials, and for each trial there was 1000ms of stimulus presentation followed by presentation of a fixation cross during a jittered interstimulus interval (ISI) between 3000ms and 5000ms. The jittering sequence was determined using a MATLAB program that randomly created series of numbers between 3000 and 5000 ms in steps of 250 ms for each participant (The Mathworks, Inc., 2011). The stimuli were presented using an in-house MATLAB script (Hwang, 2012). Each run lasted approximately 7 minutes and contained 30 unambiguous Spanish trials, 30 homograph trials, and 30 unambiguous English trials. Total time in the scanner was approximately 25 minutes including structural scanning. See Figure 4 for an example trial.
MRI Acquisition. MRI images were acquired on a Siemens Magnetom Trio 3-T MRI scanner at the SLEIC center, using T2*-weighted gradient-echo EPI sequence (TE = 25 ms, TR = 2000 ms, flip angle = 90°, FoV = 240 mm, slice thickness = 3 mm, 38 slices). Participants lied supine in the scanner with headphones to muffle the noise. They viewed the stimuli through a mirror attached to the head coil, while a tightly fitting vacuum pillow immobilized their heads. For each run, the functional scanning was always preceded by 6s of dummy scans to insure tissue steady-state magnetization. High-resolution (1 × 1 × 1 mm3) anatomical images were acquired using a T1-weighted, MPRAGE 3D gradient-echo sequence.

Data Analysis

Behavioral Data Analysis.

TVIP. Performance on the TVIP was scored according to the guidelines specified in the TVIP manual (Dunn, Lugo, Padilla, and Dunn, 1986). Specifically, for each participant a baseline (8 consecutive items correct) and ceiling (6 incorrect items out of 8 consecutive items) are calculated, and the number of errors between the baseline and the ceiling item (the largest of the 8 ceiling trials) are subtracted from the ceiling item to
calculate this score. For example, a participant who established a baseline between items 90 and 97 and a ceiling between items 117 and 125 and made 10 errors between items 98 and 125 would have a raw score of 125-10, or 115. After initial analysis, we noted that many of the items on the TVIP were Spanish-English cognates. These items were removed for a secondary analysis due to concerns that the cognates were artificially inflating participants’ scores. Performance on the two versions of the test correlated significantly (Pearson correlation=.518; p<.01), and consequently the secondary analysis was used for all further analyses.

**Flanker.** Data from the flanker task were cleaned using an in-house R-script (Bogulski & Gullifer, 2011). The script deletes outliers (responses faster than 50ms or slower than 1500ms) and computes the Flanker effect, relative costs, and mixing costs. For the purposes of this study, I focus on the classic flanker effect, which is calculated by subtracting the average reaction time (RT) to congruent trials from the average RT to incongruent trials. This statistic was consequently used as a covariate in the fMRI analyses.

**Letter-number sequencing (LNS):** Performance on the LNS task was evaluated by calculating the percent correct, or accuracy, for each participant. Reaction time data was also collected for this task, but accuracy was judged to be the more meaningful measure.

**Neuroimaging Data Analysis.** Preprocessing (motion correction, slice timing, realignment, normalization, and smoothing) was performed using SPM8 (Wellcome Trust Centre for Neuroimaging, University College London;
As is common procedure, the first three scans were excluded from data processing to minimize inclusion of images not yet at a steady state. Functional images were corrected for head motion by aligning all volumes to the middle (17/38) volume using a six-parameter rigid-body transformation. Then, the realigned time-series data were normalized according to the MNI stereotactic space and spatially smoothed by a 6-mm FWHM (full width at half maximum) Gaussian kernel. We excluded individuals with greater than 3 mm of motion, which led to a final sample of 19 participants.

Each trial type (homograph, unambiguous-English, and unambiguous-Spanish) was differentiated at the individual level, and 6 t-contrasts were run for each individual: Homograph (H)>Spanish (S), S>English (E), H>E, E>S, E>H, S>H. Exploratory group level analyses were then run for each of these contrasts at the whole-brain level, and utilized an uncorrected p-value threshold of 0.001 and specified a minimum size of 5 voxels.

In contrast, the ROI analyses were run only for the contrasts of interest specified by the hypotheses (H>S; S>E). The ROIs included the middle temporal gyrus (MTG), inferior frontal gyrus (IFG), medial frontal gyrus (MeFG), middle frontal gyrus (MFG), anterior cingulate cortex (ACC) and caudate (CAD). These ROIs were selected based on previous research, as discussed in the introduction, and were analyzed using the Wake Forest University pick atlas toolbox for SPM8 (Maldjian, Laurienti, Burdette, Kraft, 2003; Maldjian, Laurienti, Burdette, 2004). The WFU pick atlas and Talaraich client (Lancaster, Summerlin, Rainey, Freitas, Fox, 1997; Lancaster et al, 1997; Lancaster et al, 2000) was used at each session to assign names to the coordinates. These analyses were performed for all participants at both Session 2 and Session 3.
At Session 2, 29 participants were included in the analysis, while at Session 3 19 participants were included in the analysis.²

At each session, performance on the behavioral measures (TVIP, LNS, and Flanker) was entered as a covariate for each contrast of interest. Additionally, the influence of time (Session 2 & 3) and trial type (measured as the individual contrasts discussed above) were assessed by running a 2x6 ANOVA in SPM8. Relationships between the ROIs of interest (MTG, ACC, IFG, MFG, MeFG, INS) were assessed using the GIMME program (Gates & Molenaar, 2012), which conducted a euSEM analysis for both the group level maps at Sessions 2 and 3, and additionally assessed individual variation from these maps at each session.

² Five of these participants did not return for the 3rd session, while the remaining five were excluded due to movement greater than 3 mm.
Chapter 3: Results

Behavioral Results

Session 1 measures (proficiency, letter-number sequencing, flanker). Results of the Session 1 measures are summarized in Table 2. Implications of this data is detailed in the fMRI results section and Chapter 4.

Table 2: Session 1 Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Accuracy</th>
<th>Reaction Time (RT)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>Proficiency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TVIP</td>
<td>59% (7%)</td>
<td></td>
<td>46%-74%</td>
</tr>
<tr>
<td>Self-Report*</td>
<td>5 (.65)</td>
<td></td>
<td>3.25-6</td>
</tr>
<tr>
<td>Letter-Number</td>
<td>59% (14%)</td>
<td>9780ms (2609 ms)</td>
<td>33%-81%</td>
</tr>
<tr>
<td>Sequencing (LNS)</td>
<td></td>
<td></td>
<td>4858ms-13969ms</td>
</tr>
<tr>
<td>Flanker**</td>
<td></td>
<td>50.1 ms (22.6 ms)</td>
<td>12.9ms-103.5ms</td>
</tr>
</tbody>
</table>

*On a 1 to 7 scale. This score represents a composite of four self-ratings of the following abilities in their second language: Reading, Writing, Listening, and Speaking. **Flanker results are reported in reference to the flanker effect, which is computed by subtracting reaction times of congruent trials from incongruent trials.

Semantic judgment task. At Session 2 the average latency for the semantic judgment task was 937.3 ms, with a pooled standard deviation of 359.7 ms. The latency ranged from 722 to 1076.5 ms. The average accuracy was 81.9%, with a pooled standard deviation of 38.4%. The accuracy scores ranged from 65.5% to 95.4%. At Session 3, the average latency for the semantic judgment task was 875.6 ms, with a pooled standard deviation of 270.4 ms. The latency ranged from 647.5 ms to 1136.9 ms. The average accuracy for the task at Session 3 was 84.29%, with a pooled standard deviation of 36.4%. The accuracy scores ranged from 75 to 92 percent correct.
Paired sample t-tests show that accuracy at session 2 was not significantly different \((t=-1.522, p=.146)\) from accuracy at session 3. Reaction times, however, were significantly different \((t=2.146, p=.047)\) between Sessions 2 and 3, such that reaction times were shorter at Session 3 than Session 2. Although these results could be interpreted as practice effects, given that at least 4 months and up to 5 months had passed between Sessions 2 and 3, it is also possible that the RT difference is indicative of an increase in proficiency.

**Lexical decision task.** The combined data from Sessions 2 and 3 was evaluated using a 2 (Session) x 3 (Word Type) repeated measures ANOVA. The results show that there was no main effect of Session on accuracy, although there was a marginal main effect of Session on reaction time \((F(1)=3.518, p=.078)\). These results suggest that any practice effects that may have resulted from seeing the same stimuli (although in a different order) at both sessions were minimal. Critically, there was a highly significant effect of trial type on accuracy \((F(2)=21.86, p<.001)\) and RT \((F(2)=7.128, p<.05)\), suggesting that participants were processing the word types differently. Additionally, there was a marginally significant interaction between Session and Word Type for accuracy \((F(2)=3.213, p=.053)\), although there was no interaction for RT.

To investigate the nature of the interaction between Session and Word Type, we conducted a simple effects analysis of Word Type at each Session. At Session 2, there was a highly significant effect \((F(2)=26.052, p<.001)\) of Word Type on accuracy, such that participants were more accurate at identifying Spanish words (97% accuracy) than English or Homograph words (90% and 88% respectively). There was also a marginally significant effect of Word Type on RT \((F(2)=2.778, p=.076)\), as Spanish words were responded to faster \((M=722 \text{ ms})\) than English \((M=768 \text{ ms})\) or Homograph \((M=785 \text{ ms})\). At Session 3 there was also a highly significant effect of trial type on accuracy \((F(2)=9.22, p=.001)\) as well as on RT \((F(2)=21.662, \ldots)\).
Specifically, participants were 96% accurate and the fastest \((M=695\text{ms})\) at identifying Spanish words, 89% accurate and the slowest \((M=750\text{ms})\) at identifying English words, and 91% accurate with an average RT of 716 ms for identifying Homographs. This change in direction of accuracy between Session 2 and 3 for the English and Homograph words is likely the cause of the interaction between Session and Word Type in the omnibus ANOVA.

Although we did not initially expect the Spanish words to be processed more easily than the English words, post-experiment interviews suggest that the complete orthographic overlap between some of the English and homograph words (e.g. pie, vine, tan) caused the participants to begin questioning the language status of the English words, especially ones that have congruent orthography with Spanish (e.g. vote, lace).

**fMRI Results**

**ANOVA.** The results of the ANOVA suggest that although there was no main effect of time, there was a significant interaction between time and contrast type, such that there was an increase in activation at Session 3 than Session 2 for the contrasts of interest \((H>S\text{ and } S>E)\) but not for contrast types \(3 (H>E), \ 4 (E>S), \text{ or } 5 (E>H)\). Interestingly, there was also a decrease in activation for some areas between Session 2 and 3, but only for homograph contrasts \((H>S, H>E)\).

An ROI analysis of the interaction between time and contrast type (F.W.E. corrected) showed activity in the left IFG (BA 47), MeFG and MFG (BA 10), as well as the right MeFG (BA 11) and MTG. Specifically, the aforementioned increase in activation at Session 3 occurred in the right MeFG and left ACC for contrast type 1 \((H>S)\). This suggests that participants were experiencing more conflict (as evidenced by activity in the ACC and MeFG, which are both
implicated in monitoring and control, see Rodriguez-Fornells et al. 2009) in response to homographs than Spanish words at Session 3. For contrast type 2 (S>E), activity increased in the left IFG. As the left IFG has been implicated in the controlled access to meaning, this suggests that participants are increasingly treating Spanish words like their native English words (Rodriguez-Fornells et al., 2009). In addition to the increases in activation between Sessions 2 and 3, there were also decreases in the following areas. For the H>S contrast activity decreased bilaterally in the MeFG, as well as in the left IFG and right MFG, MTG, and ACC. For the H>E contrast activity decreased bilaterally in the MFG and MTG as well as in the left IFG and the right MeFG. The wide scope of the decreases in activation suggests that participants are indeed beginning to utilize a more efficient network centered on traditional language processing areas.

**Region of interest (ROI) analysis.** Beginning with the general ROI analysis that does not include covariates from Session 1 as predictors, we can observe in Tables 3 and 4 that for the H>S contrast, activity in the ACC disappears between Session 2 and Session 3, to be replaced with activation in the bilateral IFG and left MTG, although the activation in the other control areas involved stayed the same or became more dispersed bilaterally. Although this pattern is not in line with our initial prediction that homographs would be processed similarly at both sessions, it is not unreasonable to interpret the more widespread activation at Session 3 as evidence that participants are more able to explicitly recognize these words as homographs, and so the increased activation in the IFG and MTG could be due to accessing the meaning of the word in both languages. Similarly, although we did not anticipate that there would be more widespread activation at Session 3 in response to Spanish words, the appearance of the MTG is within the scope of our predictions, and the continued efforts of control areas are line with the results from previous research (e.g. Abutalebi, 2008).
Table 3: ROI Analyses at Session 2 and Session 3

<table>
<thead>
<tr>
<th></th>
<th>Without Proficiency</th>
<th>Working Memory</th>
<th>Inhibitory Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H&gt;S</td>
<td>S&gt;E</td>
<td>H&gt;S</td>
</tr>
<tr>
<td><strong>A</strong> Without Covariates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefrontal:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFG</td>
<td>6.56</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MeFG</td>
<td>10.97</td>
<td>4.97</td>
<td>3.58</td>
</tr>
<tr>
<td>CAD</td>
<td>4.37</td>
<td>4.61</td>
<td>-</td>
</tr>
<tr>
<td>ACC</td>
<td>4.83</td>
<td>5.01</td>
<td>-</td>
</tr>
<tr>
<td>MTG</td>
<td>-</td>
<td>-</td>
<td>4.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Without Proficiency</th>
<th>Working Memory</th>
<th>Inhibitory Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H&gt;S</td>
<td>S&gt;E</td>
<td>H&gt;S</td>
</tr>
<tr>
<td><strong>B</strong> Without Covariates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefrontal:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFG</td>
<td>5.93</td>
<td>-</td>
<td>4.54</td>
</tr>
<tr>
<td>MFG</td>
<td>5.40</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MeFG</td>
<td>7.05</td>
<td>4.18</td>
<td>-</td>
</tr>
<tr>
<td>CAD</td>
<td>-</td>
<td>4.91</td>
<td>-</td>
</tr>
<tr>
<td>ACC</td>
<td>-</td>
<td>4.87</td>
<td>-</td>
</tr>
<tr>
<td>MTG</td>
<td>4.91</td>
<td>5.18</td>
<td>5.35</td>
</tr>
</tbody>
</table>

Note. Values represent averaged peak T score. All correlations are positive unless written in italics for A (Session 2) and B (Session 3). Only significant values are reported.
Continuing on to the ROI analyses that included proficiency as a covariate, perhaps the most interesting result is the change from a positive correlation at Session 2 in the MeFG, a control area, to negative correlations at Session 3 in the IFG, which is associated with controlled semantic retrieval, for the Homograph>Spanish contrast. The change could be interpreted as indicating an interaction between time and proficiency, such that initially more activation in control areas allows students to focus more on the L2 and consequently boost their proficiency, but once students have reached a certain level of proficiency, they do not need to rely on those areas as much to be able to access the L2. This pattern is reminiscent of Abutalebi’s (2008) hypothesis that L2 learners initially rely more on prefrontal control areas, but later converge to the pattern of activity in the L1. Although we did not measure proficiency at Sessions 2 and 3 with the TVIP, reaction times from the semantic judgment task and the lexical decision task suggest that participants did gain some proficiency between the two sessions, which would further support the above interpretation of the results. A similar pattern was observed for the correlation between activity in the MTG and proficiency in the H>S contrast. This is opposite direction from what we expected based on the hypotheses, but may be due to the participants initially treating the homographs like English words, but at Session 3 realizing that they were in fact language-ambiguous Spanish words.

In the Spanish>English contrast that included proficiency as a covariate, we are able to observe that the MTG becomes more important at Session 3, which is in line with our hypotheses, as is the diminishing importance of the ACC after Session 2. The prefrontal activity at Session 3 may suggest that participants were changing the mechanism used for inhibiting the L1, from the ACC to the MeFG. This possibility is supported by the fact that in the analysis that
does not include proficiency as a covariate, the ACC remains involved, suggesting that it is only the more proficient participants that are making this change.

Moving to a discussion of working memory, we observe a positive correlation with the IFG for the homograph>Spanish contrast at Sessions 2 and 3. The positive correlation is in the ventral IFG. Rodriguez-Fornells et al. ‘s (2009) model of L2 processing implicates ventral IFG in semantic retrieval and consequently we may interpret the results as being consistent with our prediction that participants with higher working memory are less likely to engage control areas and better able to access meaning in their L2. For the S>E results, we observe an initial correlation in control areas (MeFG and ACC), but this correlation disappears at Session 3. Although one cannot put much faith in a null result, the disappearance of the correlations at Session 3 may suggest that with greater proficiency, working memory may become less important.

Similar to what we observed for proficiency, results from the inhibitory control analyses suggest that for the Homograph>Spanish contrast, greater IC skills allow participants to initially work harder in order to obtain proficiency, but after attaining it, participants are able to utilize control areas less, even for homographs. This pattern is predicted by our hypotheses, although we did not expect it to extend to the homograph trials. For the Spanish>English contrast, we observe roughly the same pattern as we observed for the working memory analyses, such that initially positive correlations disappear at Session 3. The similarity between these results is not unexpected, as both inhibitory control and working memory can be considered facets of executive control.
Connectivity Analysis. The Session 2 group map (see Figure 5) had an excellent fit to the data for between 90 and 100 percent of the participants, depending on the measure. Specifically, the Comparative Fit Index (CFI) and Standardized Root Mean Square Residual (SRMR) evaluated the model fit as excellent for 100% of the participants’ data, while the Non-Normed Fit Index (NNFI) and Root Mean Square Error of Approximation (RMSEA) statistics suggest that the model was an excellent fit for 90% of the participants’ data. The Session 3 group map (see Figure 5) was considered an excellent fit for 100% of the participants’ data according to the CFI, NNFI, and SRMR statistics, and an excellent fit to 90% of the data according to the RMSEA statistic.

The chi-square measure suggested that the model was not an excellent fit for any of the participants’ data at either session, but this measure is flawed in that it favors complex models and is heavily dependent on sample size. Additionally, the assumptions of the chi-square statistic are typically violated when evaluating connectivity models, and are better suited to detecting effects among ROIs, rather than determining the fit of a model (Gates, 2012). Given these shortcomings, we will not report on the chi-square further.
Figure 5. Session 2 and 3 Group Maps

Blue and purple arrows are positive and negative lagged connections, and yellow and orange arrows are positive and negative contemporaneous connections, respectively. Arrow thickness denotes the strength of the connections. MTG=Middle Temporal Gyrus, CAD=Caudate, IFG=Inferior Frontal Gyrus, ACC=Anterior Cingulate Cortex, MeFG=Medial Frontal Gyrus, MFG=Middle Frontal Gyrus
**Control Network.** For the purpose of clarity, we will collapse across the IFG, MFG, and MeFG in organizing our discussion of the connectivity results between these nodes because our initial predictions focused on the prefrontal cortex more generally. We will focus on the relationship between these nodes and the ACC and CAD, as these nodes form the basis of the control network (e.g. Abutalebi and Green, 2007). For a graphical representation of these relationships, please see Figure 6.

**CAD.** At Session 2, the connections between the prefrontal nodes and the caudate were of moderate strength. Specifically, the IFG had a lagged negative influence with a beta weight of -.33, while the contemporaneous connections between these nodes were positive (bw=.46). This general pattern held at Session 3, although the connection between these nodes strengthened, such that the negative lagged influence (bw=-.48) shifted direction such that the Caudate influenced the IFG and the positive contemporaneous connection increased to a beta weight of .76.

**ACC.** At Session 2 we observe weak to moderate lagged relationships with the prefrontal cortex, specifically a weak positive influence on the MFG (bw=.31) and a moderate negative influence from the MeFG (bw=-.63). Additionally, the contemporaneous relationship between the ACC and MeFG is very strong, with a beta weight of .97. Interestingly, this relationship disappears at Session 3, and is replaced with a moderate negative relationship with the IFG (bw=-.65). The lagged relationship with the MeFG, however, remains mostly the same (bw=-.58), and the ACC’s influence on the MFG disappears and is replaced with a weak influence on the IFG (bw=.31), although that connection is only shared by 16% of the participants. In sum, it is clear that the prefrontal cortex and ACC are intricately involved in performance on the lexical decision
task, such that the MeFG moderates the ACC’s activation, while the ACC works in concert with the MFG initially, but with the IFG at Session 3.

Figure 6. Control Network atSessions 2 and 3: Blue and purple arrows are positive and negative lagged connections, and yellow and orange arrows are positive and negative contemporaneous connections, respectively. Arrow thickness denotes the strength of the connections. MTG=Middle Temporal Gyrus, CAD=Caudate, IFG=Inferior Frontal Gyrus, ACC=Anterior Cingulate Cortex, MeFG=Medial Frontal Gyrus, MFG=Middle Frontal Gyrus
**Control and meaning network interface.** The following section will focus on the relationship between the MTG and the three main parts of the control network, the prefrontal cortex, CAD and ACC, as these connections will be informative about the relationship between semantic access and control. For a graphical representation of these connections, please see Figure 7.

**Prefrontal.** Initially the MTG has a weak negative lagged influence on the MeFG (bw=-.26) and is negatively influenced by the MFG (bw=-.36), although the MFG connections becomes positive when we examine the contemporaneous connections (bw=.51). Additionally, the MTG has a relatively weak contemporaneous relationship with the IFG (bw=.22). This IFG-MTG connection disappears in the Session 3 contemporaneous map, while the MFG relationship is preserved although weakly (bw=.23). Additionally, the MTG gains a contemporaneous connection with the MeFG at Session 3, although it is also weak (bw=.12). Similarly, the lagged connections at Session 3 are weak, with a negative connection to the IFG (bw=-.11) and a positive connection to the MeFG (bw=.11), that are only shared by 32 and 37 percent of the participants, respectively. Consequently, it appears that the MTG activity for these intermediate learners is relatively weak, as would be predicted by Yokoyama et al. (2009), although it is possible that the effects of interest were obscured by the inability of the GIMME (Gates & Molenaar, 2012) program to differentiate between trial types.

**CAD.** Interestingly, the relationship between the MTG and the Caudate is absent in both the lagged map at Session 2 and the contemporaneous map at Session 3. The connections that are apparent are weak, at both Session 2 and Session 3. Specifically, the
contemporaneous connection at Session 2 has a beta weight of .16, while the lagged relationship at Session 3 has a beta weight of .16 and was only observed in 16 percent of the participants. Although this influence is small, it is interesting that the connection between these areas shift from contemporaneous to lagged, suggesting that for some of the participants, the MTG was beginning to moderate how much control is needed, perhaps suggesting that these participants were directly accessing the meaning of the Spanish words.

**ACC.** There are no direct connections between the ACC and MTG at Session 2, either lagged or contemporaneous. The only connection between these two areas is in the contemporaneous map of Session 3, where the ACC and MTG share a weak negative (bw=-.15) connection. This does not mean that there is not an influence of the ACC on the MTG, but rather that that influence is being mediated by the PFC. If we return to the PFC results or glance at Figure 5, it becomes apparent that the ACC is primarily influencing the PFC, which in turn is influencing the MTG. This pattern is consistent with the ACC’s previously observed role as a conflict monitor (Abutalebi & Green, 2007).
Figure 7. Control and meaning network interface at Sessions 2 and 3: Blue and purple arrows are positive and negative lagged connections, and yellow and orange arrows are positive and negative contemporaneous connections, respectively. Arrow thickness denotes the strength of the connections. MTG=Middle Temporal Gyrus, CAD=Caudate, IFG=Inferior Frontal Gyrus, ACC=Anterior Cingulate Cortex, MeFG=Medial Frontal Gyrus, MFG=Middle Frontal Gyrus
In summary, it appears that the relationships between the control areas (prefrontal cortex, ACC, and CAD) are generally stronger than the connections between these areas and the MTG. It is also clear, however, that the MTG is generally more involved in the network at Session 3 than Session 2, which is in line with our initial predictions.
Chapter 4: General Discussion

Developmental Changes in Semantic Access

One of the methodological innovations our study offered in comparison with past research was the ability to track the changes in how participants access their second language over the course of a few months. To this end, our analysis utilized an ANOVA of the BOLD activity in our regions of interest (ROIs) at Session 2 and Session 3 to compare activation between these two sessions. The analysis revealed a significant interaction between time and contrast type, with increases in activation at Session 3 for H>S and S>E trials, and decreases in activation for H>S and H>E contrasts. The increase in activation for H>S trials occurred in the right MeFG and left ACC, while the increase under S>E conditions was in the left IFG.

Although we did not expect to observe increases in activity in response to homographs between the two sessions, that result is not out of line with what the literature might predict. To observe dorsal MeFG activation, which is associated with decision making (Abutalebi, 2007), and ACC activation which is associated with conflict monitoring (Abutalebi, 2012) in greater amounts for homographs than for Spanish words suggests that participants were experiencing less conflict and having an easier time making decisions about the language membership of the Spanish unambiguous words at Session 3, which is in line with our predictions. Additionally, although Van Heuven et al. (2008) observed inferior PFC activation in response to homographs, it is not unusual to also observe activation in this area (specifically in this case the left IFG) in response to unambiguous second language words (Abutalebi, 2007).

Similarly, the decreases in activation between Session 2 and Session 3 for the H>S and H>E trials were also concurrent with our predictions, which were that at Session 3 the
participants should be utilizing more efficient, L1-like networks. Specifically, the decreases occurred in the right ventral MeFG, left IFG, right MFG, right MTG, left MeFG, and right ACC for the H>S trials, and bilaterally in the MFG, SFG, and MTG, as well as in the right rectal gyrus (part of the ventro-medial PFC), right MeFG, left precentral gyrus and left IFG for the H>E trials. These areas are predominantly control areas, commonly associated with conflict monitoring (ACC), decision-making (MeFG), cognitive control (SFG, IFG, MFG), and motor control (precentral gyrus) (Abutalebi, 2007; Guo, Liu, Misra, Kroll, 2011). This pattern supports Abutalebi and Green’s (2007) convergence hypothesis, which suggests that as learners become more proficient, they will utilize control areas less.

The decrease in activation in the MTG for the homographs in the H>S and H>E trials is more difficult to explain, but may be due to participants’ growing knowledge of the L2 and consequent greater ability to recognize the words as homographs. This fits with the observed increase in activation for these trials in the MeFG and ACC, suggesting that the participants were experiencing more conflict in processing the homographs at Session 3. Unfortunately, we did not question the participants specifically on their knowledge of the homograph status of the words after the experiment, so it is difficult to verify this hypothesis.

In addition to the ANOVA, we also utilized Group Iterative Multiple Model Estimation (GIMME; Gates & Molenaar, 2012) to examine the difference in the group networks between sessions. This type of analysis is able to examine lagged effects (the effect of one ROI at time one on a second ROI at time two) as well as contemporaneous effects (the relationship between two or more ROIs at time one). Because lagged effects give information about directionality as well as the strength of the relationship (expressed in terms of beta-weights), we will begin by discussing the changes in the lagged effects between Sessions 2 and 3. There is an overall trend
to observe a change from conflict-first processing to meaning-first processing, as shown by the increased involvement of the IFG and MTG at Session 3. Specific examples include the MeFG losing its influence on the MTG after Session 2, the IFG’s gained influence at Session 3 on the ACC and the Caudate, and the MTG’s gained influence over the IFG as well as the MeFG at Session 3.

Unlike lagged effects, contemporaneous effects cannot be interpreted with directionality. The pattern we observed for the lagged effects, however, does appear to also hold for the contemporaneous effects. Specifically, we observed that control areas such as the MFG and Caudate lost connections between Session 2 and 3, while areas associated with semantic processing such as the IFG and MTG gained connections. Although there is no change in the overall size of the network as we initially predicted, the shift to an IFG and MTG centric network does fall in line with the spirit of those predictions, which was that participants should be utilizing these areas more and the control areas less.

**Semantic Access in the L1 and the L2**

Moving from the more general changes between sessions that informed us about the developmental trajectory of semantic access in these L2 learners, we will now examine the differences between trial types to inform us about the primary goal of the experiment: how semantic access in the L2 compares with L1 semantic access. At Session 2, participants utilized the left MeFG and bilateral ACC—areas associated with decision making and conflict monitoring, respectively—more when processing unambiguous Spanish words compared with L1 words. This suggests that participants were not directly accessing the meaning of these words, but instead first had to inhibit the first language, as predicted. Interestingly, there was also some overlap with the E>S contrast, such that the left MeFG was active for both contrasts, although
the E>S pattern was more dorsal. This overlap could be interpreted as the beginning of convergence between the L1 and L2 patterns, as would be predicted by Abutalebi and Green (2008). At Session 3, participants continued to utilize the MeFG and ACC, but one can also observe increased use of the MFG, caudate, and MTG for Spanish words compared to English words. Although MFG is not usually explicitly referred to in the language literature, it overlaps with the dorso-lateral prefrontal cortex (DLPFC), which is often discussed in relation to its role in mediating conflict in language processing (e.g. Guo, Liu, Misra & Kroll, 2011). This pattern of activation suggests that these participants are still experiencing conflict, which is not surprising given the nature of the task, but also increasing automaticity in the access of L2 meaning, as evidenced by the increased activation in the MTG. This interpretation is partially confirmed by the results of the E>S contrast, which overlap with the S>E results in the right MTG and left cingulate cortex, suggesting that at Session 3 participants’ L2 processing patterns are converging with their L1.

Focusing specifically to the MTG, the results of the current study confirmed some previous research findings (Abutalebi & Green, 2008) and differed from others (Yokoyama, 2009). Specifically, the current finding that with increasing experience—as measured here by the differences between Sessions 2 and 3—with the language the MTG becomes more active for the L2 corresponds with Abutalebi and Green’s (2008) convergence hypothesis. It also suggests that the learners Yokoyama (2009) tested may not have been proficient enough in their L2 to observe this MTG activation.

One limitation to this claim, however, is that it appears that the learners were processing the L1 words and homographs similarly. This is evident in the behavioral scanner data from the lexical decision task, where Spanish words were processed faster and more accurately than either
the homograph or English words. This may suggest that the MTG activation we observe at Session 3 for the homograph contrast is merely an artifact of accessing the English meaning first. This does not diminish the finding that the MTG activation increased between sessions for the unambiguous Spanish words, suggesting that the learners were in fact beginning to process the L2 more similarly to their L1.

**Influence of Individual Differences on L2 Semantic Access**

The second goal of this study was to investigate how the linguistic processing of L2 words may be influenced by individual differences in linguistic (L2 proficiency) and non-linguistic (working memory and inhibitory control) domains. Given that none of these factors correlated with each other in the behavioral results, each factor will be examined separately below.

**Proficiency.** The proficiency results vary greatly by contrast, and consequently the Homograph>Spanish and Spanish>English contrasts will be discussed separately. At Session 2, proficiency was positively correlated with activity in the MeFG and MTG for the Homograph>Spanish contrast. At Session 3 the IFG and MTG are negatively correlated with proficiency. This may suggest that initially participants were treating the homographs like English words, and automatically retrieving the English meaning, but by Session 3, the more proficient learners were able to identify that these were Spanish homographs and inhibit the English translation, while the less proficient learners continued to access the English meaning.

For the Spanish>English contrast, proficiency was initially positively correlated with ACC activity while at Session 3, this activity shifted so that proficiency was positively correlated with the left MeFG, as well as the left and right MTG. This pattern of results suggests that at
Session 2, high proficient learners were also the learners who were trying harder, as evidenced by the positive ACC activity. By Session 3, however, these learners were able to adopt a more native-like activation pattern, with highly proficient learners shifting focus to the MTG and MeFG as compared with the ACC. We propose in the results section that this shift in areas may represent a shift in the processing mechanism for more proficient learners, but why more proficient learners would use the MeFG as opposed to the ACC is unclear and should be subject to future research.

**Working Memory.** Of the three covariates, working memory performance correlated with the fewest regions of interest, and 75% of those correlations were only present at Session 2, while the correlation that remained at Session 3 was in response to homographs. This suggests that working memory is primarily important at early stages of learning, when it is most difficult to inhibit the L1 and consequently the learner has to juggle the highly activated representations in both languages. At later stages, although it has been shown that even highly proficient bilinguals activate both languages in most scenarios, it may be that the need to inhibit is less conscious, and consequently draws less on the explicit manipulation processes that are measured by the letter-number sequencing task (Marian, Spivey, and Hirsch, 2003; Hartsuiker, Costa, and Finkebeiner, 2008). An interesting scenario for future research would be to examine if the type of conflict predicted by the initial stages of the RHM, when learners are explicitly translating their L1 thoughts (or even assignments, as many foreign language teachers can attest) instead of attempting to think in the L2, correlates more with working memory than inhibitory control measures (Kroll & Stewart, 1994).

**Inhibitory control.** The relationship between inhibitory control performance and BOLD activity largely fell in line with our predictions, as at Session 2 inhibitory control performance
was positively correlated with activity in the left IFG and MeFG while at Session 3 inhibitory control performance was negatively correlated with activity in those areas. This suggests that learners with higher performance on inhibitory control measures are able to utilize more efficient networks (i.e. networks with lower levels of activation). While we did not predict that the correlations would be initially positive, we expect that this phenomenon is related to the fact that learners may not have initially recognized all the homographs for what they were. An alternate explanation could be that learners need to develop a baseline level of proficiency before differences in inhibitory control can be observed. For the Spanish contrast, however, activity increased bilaterally in the MeFG, left MTG, and right MFG with inhibitory control performance at Session 2 but these correlations dropped out at Session 3. This pattern of results does not support the baseline proficiency hypothesis but instead suggests that inhibitory control is less important once learners are able to directly access L2 meanings, similar to what we observed with the working memory results and similar to what has been observed previously by Abutalebi (2008).

Although these results are promising, it is important to note several limitations of this research. Perhaps the largest was a flaw in the design, where we did not collect TVIP scores at Sessions 2 and 3, and this hinders our ability to claim that the differences in these sessions are truly due to increases in proficiency. However, we are able to observe decreased reaction times in the lexical decision task and semantic judgment task between sessions, which is indicative of increased proficiency. Additionally, as noted earlier in this chapter, it would have been helpful to collect information not only on participants’ familiarity with the stimuli from the lexical decision task, but also their familiarity with the concept of a homograph and their ability to recognize one.
Even given these limitations, the results from this study have largely supported our initial predictions, and consequently provide further support for Green’s (2003) convergence model of second language processing, as well as Hernandez and Li’s (2007) integration model. Future work should seek to differentiate these hypotheses by analyzing L2 learners’ comprehension of sentences as well as individual words, as this is where the two hypotheses diverge. Green’s (2003) model would suggest that given sufficient proficiency L2 learners will eventually process sentences similarly to L1 speakers, while Hernandez and Li (2007) suggest that even with increased proficiency learners will continue to process syntactic structures in their L2 differently from how they would in their L1.

Additionally, the results of this study provide new evidence in light of the RHM, which suggests that learners must initially translate words into their L1 in order to access meaning. Our results have shown that for unambiguous L2 words skills associated with translation, such as inhibitory control and working memory, are largely non-influential with increased experience. Additionally, activation of areas associated with semantic processing in response to unambiguous L2 words occurred only at Session 3, unless the participant had high inhibitory control skills.

The theoretical ramifications of this experiment, and its ability to inform us about the nature of the neurological instantiation of the differences between L1 and L2 processing, rely primarily on its ability to inform pre-existing theories (e.g. the RHM and the convergence hypothesis; Green, 2003; Kroll & Stewart, 1994) of the consequences of individual differences in working memory and inhibitory control skills. Our results suggest that one’s success as a bilingual, as measured by the ability to process the L2 in a manner similar to the L1, may be pre-
determined by one’s pre-existing abilities. Future research should examine the scope and time
course of this relationship.
References

*Acta psychologica, 128*(3), 466–78.

*Cerebral cortex, 22*(9), 2076–2086. doi:10.1093/cercor/bhr287


*Psychological Science in the Public Interest, 10*(3), 89–129.


