Brain and Language 197 (2019) 104661

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

A longitudinal investigation of structural brain changes during second language learning

Jennifer Legault[a,b,*], Angela Grant[a,c], Shin-Yi Fang[a], Ping Li[a,*]

[a] Department of Psychology and Center for Brain, Behavior, and Cognition, Pennsylvania State University, United States
[b] Department of Linguistics and Cognitive Science, University of Delaware, United States
[c] Department of Psychology, Missouri Western State University, United States

ARTICLE INFO

Keywords:
Second language learning
Structural magnetic resonance imaging
Individual differences

ABSTRACT

Few studies have examined the time course of second language (L2) induced neuroplasticity or how individual differences may be associated with brain changes. The current longitudinal structural magnetic resonance imaging study examined changes in cortical thickness (CT) and gray matter volume (GMV) across two semesters of L2 Spanish classroom learning. Learners' lexical processing was assessed via a language decision task containing English and Spanish words. Our findings indicated that (1) CT increased in the left anterior cingulate cortex (ACC) and right middle temporal gyrus (MTG) after L2 learning, (2) CT in the right MTG increased in individuals who were better able to discriminate between native language and L2 words, and (3) CT in the left ACC was correlated with functional connectivity between the ACC and MTG. These findings indicate that L2 lexical development is associated with functional and structural changes in brain regions important for cognitive control and semantic processing.

1. Introduction

In recent years a large number of studies have suggested that second language (L2) experience leads to changes in language and cognitive control functional networks and brain structures (Li, Legault, & Litcofsky, 2014; Lövdén, Wenger, Mårtensson, Lindenberger, & Bäckman, 2013; Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009). However, many studies in this domain are cross-sectional studies that compare bilinguals with monolinguals, and very few have tracked the longitudinal effects of L2 learning and the individual differences in performance that might correlate with these brain changes. To better understand how these neural regions cooperate with one another to allow for L2 learning and L2 processing, an increasing number of studies have adopted a functional connectivity approach to examining these complex relationships. The current study builds upon a previous functional connectivity study of L2 learning (Grant, Fang, & Li, 2015) by examining the relationship between structural brain changes and individual differences in linguistic ability and experience. Furthermore, we examine the relationship between longitudinal structural brain changes and functional connectivity changes based on L2 learning experience, aiming for a multimodal understanding of experience-dependent neuroplasticity.

A distinct advantage to using longitudinal structural magnetic resonance imaging (sMRI) approaches is that they are uniquely positioned to infer causal relationships between L2 experience and subsequent brain structure changes (Li & Grant, 2015). These brain structure measures are posited to reflect aggregate changes in axonal architecture, dendritic branches, and various aspects of synaptic changes, including synaptogenesis (Zatorre, Fields, & Johansen-Berg, 2012). In contrast, functional magnetic resonance imaging (fMRI) approaches examine levels of blood oxygenation and deoxygenation as a proxy measure for neural activity, and their interpretations may vary by fMRI task used. Since both sMRI and fMRI approaches provide distinct and valuable insight into neuroplasticity, multimodal approaches which employ both measures are particularly well-suited to provide a nuanced view of the relationships between brain structure, neural activity, and behavioral performance. In the current study's novel comparison of brain structure changes with task-based functional connectivity changes, our findings demonstrate that regions that gain functional connections as a function of L2 experience also increase in cortical thickness (CT), particularly highlighting the role of the anterior cingulate cortex (ACC) and middle temporal gyrus (MTG) in intermediate L2 learning-based neuroplasticity. Further, changes in CT and GMV varied based on individual differences in native language (L1) and L2

---

* Corresponding authors at: 800 Barksdale Rd. Room 103B, Newark, DE 19711, United States (J. Legault). 452 Moore Building, University, PA 16801, United States (P. Li).

E-mail addresses: jegault@udel.edu (J. Legault), pul8@psu.edu (P. Li).

https://doi.org/10.1016/j.bandl.2019.104661
Received 30 October 2018; Received in revised form 15 July 2019; Accepted 16 July 2019

performance and experience. In the following sections, we detail previous findings regarding neural functions of lexical processing in L1 and L2 and emphasize the importance of longitudinal multimodal approaches and individual difference analyses in examining whether experience-dependent brain changes may be related with other cognitive factors.

1.1. Neural substrates of lexical processing in L1 and L2

When students are first beginning to learn a second language, a major step includes vocabulary learning, which involves mapping a new L2 lexicon onto pre-existing mental concepts, which are already connected to words in their L1. Examining how these learners process this lexical information in their L1 and L2 at different points in learning can help to resolve conflicting theories of how adult learners integrate this information.

In a previous study, Grant et al. (2015) used functional connectivity analyses to address the predictions of two diverging theories of L1 and L2 lexical processing in adult L2 learners. Specifically, the developmental variant of the Bilingual Interaction Activation model (BIA-d; Grainger, Midgley, & Holcomb, 2010) and Convergence Hypothesis (CH; Green, 2003; Green, Crinion, & Price, 2006) – make competing predictions on the timescale and extent of cognitive control and on the divergent patterns of L1 and L2 lexical processing in early versus late stages of L2 learning. Within our study's context, cognitive control is defined as the mental ability to store and process new information in a flexible manner. This ability includes both inhibitory control (in particular, inhibiting a non-selected language or item) and working memory (the ability to store short-term memory items and perform some mental calculation or translation to the information). The BIA-d model predicts that semantic processing in the L2 will be initially processed through the L1 rather than having a direct connection to the shared semantic store during early learning stages, but will inhibit these L1 connections in order to facilitate direct connections to the semantic store during later stages of L2 learning. By contrast, in the early L2 learning stage the CH model predicts greater activation in cognitive control regions for L2 versus L1 lexical processing, with decreasing dependency on cognitive control and re-focused resources on a language – network during proficient, late stage L2 learning (see Fig. 1). Notably, the Grant et al. (2015) study used a longitudinal functional connectivity design to address the developmental time course in the use of these language and cognitive control networks across two semesters of L2 learning. Their study found that adult L2 Spanish learners relied on language and cognitive control networks for the early stages of L2 learning and switched to relying more on semantic processing regions during later learning stages, in support of the CH model.

While Grant et al. focused on examining the functional connectivity changes across late L2 learning, the current study adds to our knowledge of L2 experience-dependent neuroplasticity by examining the changes in brain structure that accompany L2 learning. Specifically, the current study examines, in the same participants, CT and GMV changes across L2 learning in the functionally connected regions identified by Grant et al., in order to gain a more comprehensive understanding of potentially causal relationships between changes in brain structure and L2 experience. Previous fMRI work has identified a number of key regions of interest (ROIs) involved in bilingual language processing, along with increasing evidence from sMRI studies that are consistent with the fMRI-based ROIs (see Li et al., 2014; Pliatsikas, in press; Stein, Winkler, Kaiser, & Dierks, 2014 for reviews). Importantly, these key regions have previously demonstrated both functional and structural changes in response to L2 experience. Specifically, semantic and lexical processing regions include the bilateral middle temporal gyrus (MTG) and inferior frontal gyrus (IFG). The MTG has been shown to increase in GMV with L2 learning (Mårtensson et al., 2012) and is posited to be a semantic processing 'hub' required for processing of lexical and conceptual semantic information (Binder & Desai, 2011; Hernandez, Woods, & Bradley, 2015; Rodríguez-Fornells et al., 2009). The IFG is another hub-like region, and has been shown to increase in both CT and GMV in response to L2 learning (Hosoda, Tanaka, Nariai, Honda, & Hanakawa, 2013; Klein, Mok, Chen, & Watkins, 2014; Mårtensson et al., 2012; Stein et al., 2012). The IFG plays a major role in semantic retrieval and selecting relevant semantic information over distractors (Rodríguez-Fornells et al., 2009; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997) and is consistently activated during various aspects of L2 processing and can be further sub-divided depending on the language tasks (Abutalebi, 2008; Parker Jones et al., 2012).

In addition to the semantic network, studies have also identified a number of key language control network ROIs, which include (1) the bilateral anterior cingulate gyrus (ACC), (2) the middle frontal gyrus (MFG), and (3) the caudate nucleus (CN). The ACC has been shown to

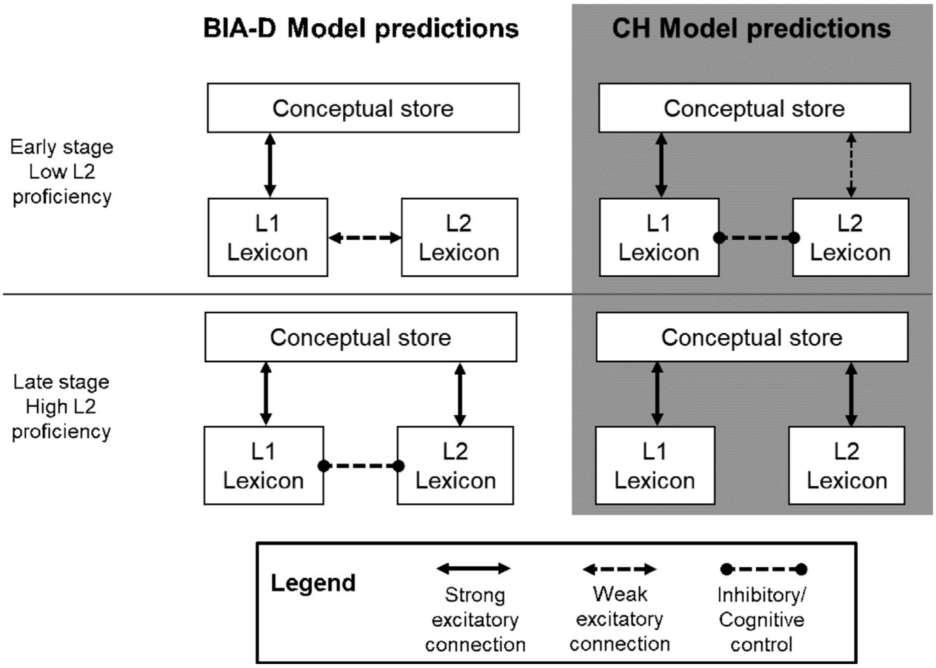

Fig. 1. Generalized depiction of predictions for early versus late stage L2 learning according to the Bilingual Interactive Activation developmental (BIA-D) model versus the Convergence Hypothesis (CH) model. The BIA-D model predicts that only late stage or highly proficient learners will use inhibitory/cognitive control, while the CH model predicts that early stage learners will use inhibitory/cognitive control, which will decrease with increasing proficiency.

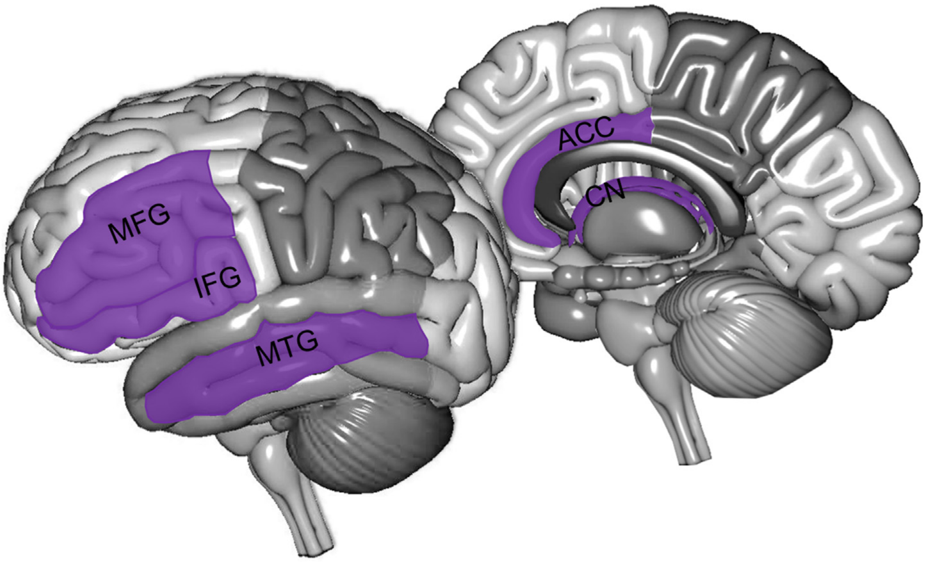

Fig. 2. Depiction of the regions of interest (ROIs) included in the current study. Our ROIs include the anterior cingulate cortex (ACC), caudate nucleus (CN) inferior frontal gyrus (IFG), middle frontal gyrus (MFG), and middle temporal gyrus (MTG).

be important in L1 and L2 conflict monitoring, where increased conflict monitoring ability is associated with decreased functional activity and increased GMV in the ACC (Abutalebi et al., 2012). The MFG has been observed to increase in CT with L2 learning (Mårtensson et al., 2012) and is posited to be involved with inductive reasoning (Rodríguez-Fornells et al., 2009) and word retrieval (Price, 2010). Finally, the CN has been shown to (a) be larger in bimodal bilingual versus monolingual participants (Zou, Ding, Abutalebi, Shu, & Peng, 2012), (b) change in structural connectivity with frontal regions with L2 experience (Van de Putte et al., 2018), (c) change in shape in response to bilingual immersion (DeLuca, Rothman, & Pliatsikas, 2018), (d) correlate with nonverbal cognitive control (Abutalebi et al., 2012), and (e) correlate both structurally and functionally with phonemic fluency (Grogan, Green, Ali, Crinion, & Price, 2009). For a visualization of all our ROIs, please see Fig. 2 below.

Importantly, few of the studies mentioned above have adopted the multimodal approach necessary for a comprehensive view of experience-dependent neuroplasticity. Specifically, Abutalebi et al. (2012) found that bilinguals who used their ACC more efficiently (as exhibited by less functional activity in this region during a flanker task) and performed with higher accuracy as compared to monolinguals, also showed greater GMV in this region. Furthermore, they found a positive correlation between GMV in the ACC and the fMRI conflict effect for the ACC in bilinguals, corroborating the important role of the ACC in conflict monitoring in bilinguals.

Additionally, Zou et al. (2012) examined functional activity and GMV in bimodal bilinguals versus monolinguals, finding that GMV in the left CN positively correlated with functional activity in this region during a language switching task for bilinguals. Both of these studies observed a convergence of their behavioral, functional, and structural findings, highlighting the importance of multimodal approaches in cross-validating the relationship between neural substrates and language performance. However, both of these studies focused on comparing bilinguals to monolinguals, at only one time point. So far, no study has used a longitudinal multimodal approach to examine the neurodevelopmental relationships between bilingual lexical processing, neural structure, and functional connectivity across L2 learning stages. The current study begins to address this gap in the literature.

1.2. Individual difference relationships with neuroplasticity

An increasing amount of research has recently focused on examining how individual differences in performance might be associated with various aspects of L2 experience and neuroplasticity. The bilingual brain literature has so far focused on several factors in L2 experience that are key to modulating functional and structural brain responses (see Hernandez & Li, 2007; Hernandez, 2013 for reviews): (1) age of acquisition (AoA) for the L2, which describes the age at which a participant first learned a new language, (2) L2 proficiency, a measure of how well a person can perform in a second language across domains such as reading, writing, speaking, and listening, (3) L1 proficiency, which likewise describes how well a person can perform in their native language, and (4) cognitive control abilities, such as inhibitory control and working memory.

A number of functional connectivity studies have indicated that individual differences in L2 performance can modulate the functional connectivity networks, such that successful L2 learners showed different network connectivity as compared to less successful learners. For example, participants who were more effective L2 learners showed additional functional connections to the IFG, MFG, and IPL as compared to less successful L2 learners after six weeks of L2 learning (Yang, Gates, Molenaar, & Li, 2015). Further, successful learners of auditory pitch discrimination showed greater global functional connectivity in a fronto-temporal network as compared to less successful L2 learners (Sheppard, Wang, & Wong, 2012). Grant et al. (2015) also found that participants with greater inhibitory control (as measured by a smaller conflict effect on the Flanker task) no longer relied on activating an extensive network to process L2 words at the end of L2 learning as compared to less efficient processing by participants showing greater conflict effects. Furthermore, Chai et al. (2016) have shown that resting state functional connectivity in regions such as the ACC correlated with improvement in lexical retrieval of L2 items during spontaneous speech after 12 weeks of intensive language training. These findings further indicate cooperation between cognitive and language control regions during L2 learning.

Structural connectivity studies using diffusion tensor imaging (DTI) have shown that white matter microstructure measures vary based on L2 proficiency and AoA. For example, Xiang et al. (2015) have found laterality effects varied based on L2 learning stage and subsequently L2 proficiency level. Furthermore, studies have found that WM microstructure in tracts that connect language regions such as the IFG and superior temporal gyrus (STG) tend to increase with L2 experience or with AoA (Hofstetter, Friedmann, & Assaf, 2017; Mamiya, Richards, Coe, Eichler, & Kuhl, 2016; Luo et al., 2019; Nichols & Joanisse, 2016).

Impressively, these changes have been found to correlate with lexical learning rate and to occur on a rapid timescale, even over the course of one hour of L2 training (Hofstetter et al., 2017).

Importantly, DTI studies tend to corroborate fMRI findings, with several studies finding a correlation between white matter microstructure and functional connectivity or activity (Mollink et al., 2019; Van de Putte et al., 2018). Van de Putte et al. (2018) tracked functional activity and white matter (WM) microstructure changes in students training to be interpreters versus translator controls, finding increased WM connectivity across regions that were functionally active, which correlated with language and cognitive control ability in translators as compared to controls. Together, these findings suggest that intense language experience recruits various language and cognitive control processes.

Recent sMRI studies examining individual differences related to brain structure show relationships that are consistent with fMRI studies (Li et al., 2014), and these relationships are dependent on the key variables mentioned above, namely AoA, L1 proficiency, L2 proficiency, and cognitive control ability. For example, greater CT and GMV in language and cognitive control regions have been found to be associated with increased L1 and L2 proficiency (Hosoda et al., 2013; Mechelli et al., 2004; Stein et al., 2012), earlier AoA (Grogan et al., 2012; Mechelli et al., 2004), decreased conflict effect (Della Rosa et al., 2013) and increased conflict monitoring ability (Abutalebi et al., 2012). Furthermore, DeLuca et al. (2018) examined GMV and WM in adults learning an L2 in an immersive environment. They found GMV increases in the cerebellum commensurate with the length of immersion and the age of acquisition of the L2. Likewise, Berken, Gracco, Chen, and Klein (2016) examined highly proficient sequential versus simultaneous bilinguals, indicating greater GMV in the left putamen in simultaneous L2 learners, as compared to sequential bilinguals. Another recent longitudinal study conducted by Hervais-Adelman, Moser-Mercer, Murray, and Golestani (2017) examined cortical thickness changes in interpretation trainees as compared to multilingual controls, finding increases for the interpretation trainees in regions implicated with various functions of both language processing and cognitive control. These findings are in line with current theories that L2 processing may recruit cognitive control resources. However, little research has been conducted to examine how multiple aspects of L1, L2, and cognitive ability performance is associated with longitudinal changes in brain structure. The current study aims to systematically examine how individual differences in AoA, language proficiency, and cognitive control may be associated with L2 experience-dependent changes in CT and GMV across two semesters of intermediate-level Spanish classroom learning.

1.3. Gray matter measures

Structural MRI allows for non-invasive examination of gray matter (GM), which includes the cell bodies and dendrites of neurons as well as synapses and glial cells. Two key measures of GM that can be derived from sMRI scans include CT and GMV, and they have both been previously used in studies of bilingual processing. We used Freesurfer software to preprocess and analyze all structural data given its ability to accurately measure and model longitudinal data across scanners and protocols (Reuter, Schmansky, Rosas, & Fischl, 2012). Freesurfer's calculations of both CT and GMV have been cross-validated with histological analyses (Rosas et al., 2002) and are shown to have a high test-retest reliability (Reuter et al., 2012). The current study uses both CT and GMV measures, because previous studies have shown that together, these two measures can provide a comprehensive picture of structural neuroplasticity across time (Lemaitre et al., 2012; Narr et al., 2005). While previous studies have found that CT networks map onto resting-state functional connectivity networks (Chen, He, Rosa-Neto, Germann, & Evans, 2008; He, Chen, & Evans, 2007), the current study uniquely examines the relationship between CT and task-based functional connectivity networks.

A number of differences between the CT and GMV measures should be noted. Using Freesurfer, CT is measured as the distance between the gray and white matter boundaries along the surface of the brain, whereas GMV is calculated as the product of CT and the surface area (Fischl & Dale, 2000). CT has been shown to be a more sensitive measure as compared to GMV (Lemaitre et al., 2012) and has been posited to be more closely associated with individual differences in cognitive ability (Narr et al., 2007). However, the majority of studies examining brain structure and cognition have focused on GMV as a measure, and therefore GMV provides a rich literature to draw from (Li et al., 2014). Further, while both CT and GMV are posited to reflect neural remodeling such as axon sprouting, dendritic branching, and synaptogenesis (Zatorre et al., 2012), due to the curved nature of the outer surface of the brain these two measures are sometimes at odds with one another (Chung, Dalton, Shen, Evans, & Davidson, 2006). This is because the density/volume of the outer surface is not uniform along the gyri and sulci of the brain, and the outer surface is folded in differing degrees across brain regions and people. Thus, some regions with more cortical folding will be thinner (lower CT) and yet may register as having greater GMV (Chung et al., 2006). To take into accounts any effects of folding or sensitivity, we therefore include both measures of gray matter structure in our analyses.

1.4. The current study

Previous studies as reviewed above have shown learning-dependent reorganization in functional activity, functional connectivity, and structural brain patterns in response to L2 experience, especially in regions implicated in a language and cognitive control networks. The current study examines the relationship between CT and GMV changes longitudinally with individual differences in language and cognitive ability performance across L2 learners. Our study makes two main predictions: (1) the L2 learners, as compared to non-learning controls, will show increased CT and/or GMV in regions previously shown to gain or make new functional connections based on fMRI responses (Grant et al., 2015), and (2) CT and GMV changes for L2 learners in these regions will vary as a function of individual differences in performance in relevant tasks of language and cognitive ability. For example, we expect that flanker performance will be associated with changes in the ACC and CN, while L2 proficiency should most likely be associated with changes in the MTG and IFG. To identify the correspondence between fMRI and sMRI measures, we additionally performed a novel investigation into whether CT experience-dependent changes correlated with task-based functional connectivity. Examination of these relationships between brain structure, functional connectivity, and individual differences in performance will provide a multimodal understanding and a potentially causal view of the timescale and nature of L2 learning-dependent neuroplasticity.

2. Methods

2.1. Participants

The L2 learning group comprised 24 native English speakers (mean age = 20.58) who were registered for intermediate level Spanish courses at the Pennsylvania State University (see Grant et al., 2015). All of the participants were classroom learners of Spanish as their second language, and had started learning Spanish on average at age 12.25 (SD = 2.44). Six of the participants reported experience abroad, from 1.5 to 6 months, with an average of 3.42 months. Four participants reported knowledge of an additional language, including Arabic, French, and German. All participants were right-handed and naïve to the purpose of the experiment.

Recent studies have encouraged the use of data-sharing consortiums due to advantages such as the ability to provide a consistent control across various studies and their ability to supplement studies that are not able to acquire funding for additional participants, considering the prohibitive costs of MRI scanning. For these reasons, we have used a control group from the Consortium for Reliability and Reproducibility (CORR; Zuo et al., 2014). This control group consisted of 20 participants (mean age: 21.95; *SD* = 1.35; age range: 18–26) who were part of a healthy control test–retest experiment at the Mind Research Network, New Mexico, as part of the CORR that provides MRI data-sharing with the research community (Zuo et al., 2014). Unfortunately, this group did not provide any information on language experience or proficiency of these individuals. Informed consent was obtained from all individual participants included in the study, and all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

2.2. Measures of language performance

2.2.1. Language Decision (LD) task

This task was used as a measure of Spanish (L2) performance and the ability to distinguish English (L1) from Spanish words. Participants were presented with a written word (in either Spanish, English, or ambiguous Spanish-English homographs) in the middle of the screen, and were instructed to indicate whether the word was a Spanish word or not via button press (1 for Spanish words, 2 for non-Spanish words). There were 3 runs which each included 30 unambiguously Spanish word trials, 30 unambiguously English word trials, and 30 Spanish-English homograph word trials, for a total of 90 trials per condition. Homographs were written words that shared similar orthography in Spanish and English but had different meanings (e.g. “pie” means foot in Spanish whereas it refers to a dessert in English), and were instructed to be counted only as Spanish words in the LD task. Each trial included stimulus presentation for 1000 ms followed by a jittered inter-stimulus interval (ISI) ranging between 3000 and 5000 ms in steps of 250 ms (further details of experimental tasks and behavioral measures are provided in Grant et al., 2015).

2.2.2. Spanish semantic judgement task

This task was used as a measure of Spanish (L2) performance. Participants saw a word presented in the middle of the screen and were asked to judge whether the item referred to a living or non-living thing. Stimuli consisted of 87 Spanish nouns that were not used in the LD task.

2.2.3. Test de Vocabulario en Imagenes Peabody (TVIP)

This task is the official Spanish translation of the Peabody Picture Vocabulary Task and was used as a measure of Spanish (L2) proficiency (Dunn, Padilla, Lugo, & Dunn, 1986). Participants heard a Spanish word and were asked to indicate which of the four pictures on the screen corresponded to the auditory word. Forty-one items that were cognates (words that have a shared meaning and spelling in both Spanish and English) were removed from the analyses, due to possible inflation of scores, leaving a total of 84 trials in the analyses.

2.2.4. Language history questionnaire

The LHQ (Li, Sepanski, & Zhao, 2006) is a comprehensive survey on language history and experience. This online tool assesses the participant’s background in any of their used and/or learned languages, including their self-reported age of acquisition (AoA) for Spanish. This AoA measure was used as a variable of interest in our analyses.

2.3. Measures of cognitive ability

2.3.1. Flanker task

This task was used as a measure of inhibitory control (adapted from Emmorey, Luk, Pyers, & Bialystok, 2008). Participants were presented with five arrows on the screen and were asked to indicate the location of the middle (3rd) arrow. There were three trial types used: congruent trials (where all arrows were pointing in the same direction), incongruent trials (where the middle arrow was in the opposite direction as the other arrows) and neutral trials (where the surrounding arrows had no arrowheads and therefore were just lines). Any outliers (defined as responses that were slower than 1500 ms or faster than 50 ms) were removed from our computations, where an average of 2% of trials were removed. The variable of interest used in our analyses is the traditional flanker effect, calculated by the average reaction time in the incongruent trials minus the average reaction time in the congruent trials.

2.3.2. Letter number sequencing task

This task was used as a measure of phonological working memory (Wechsler, 1997). Participants heard a sequence of letters and numbers (e.g. f8c1) and were instructed to re-order these stimuli in ascending numeric and alphabetic order (e.g. 18cf). These sequences increased in task difficulty starting with strings of 2 characters up to strings of 9 characters, and the participants were instructed to type in their responses on the computer with no time limit. We included accuracy performance as a predictor in our individual difference regressions.

2.4. Procedure

Participants in the L2 learning group first completed a cognitive testing session that assessed participant’s L2 AoA and proficiency, as well as their working memory and executive function abilities. These cognitive tasks (as described above) included the LHQ to measure the AoA, the TVIP to measure their L2 proficiency, the LNS task to measure their phonological working memory, and the Flanker task to examine their inhibitory control ability. Next, participants underwent their first MRI scanning session where participants performed the Spanish Semantic Judgment (SSJ) task in the mock scanner and the Language Decision (LD) task in the MR scanner. Four months later, participants completed an MRI scanning session where they underwent the SSJ and LD tasks for a second time inside the MRI scanner. All sessions were completed during the academic semester while participants were actively learning and using their second language. Participants in the control group only underwent sMRI scanning sessions four months apart, with no language learning in between the two scans.

2.5. SMRI acquisition

MRI structural images for the L2 learners were acquired using a Siemens Magnetom Trio 3T at Pennsylvania State University (in the Social, Life, and Engineering Sciences Imaging Center). Participants were first habituated to the scanning environment by being placed in a mock scanner where they performed an SSJ task. Participants then entered the scanner where stimuli were projected onto a rear screen, and viewed the stimuli in a mirror mounted on the head coil. T1-weighted anatomical images were collected with acceleration factors on (IPAT = 2 GRAPPA), in interleaved order (160 slices, repetition time [TR] = 1650 ms, echo time [TE] = 2.03 ms, field of view [FOV] = 256 × 256 mm, with voxel size = 1 × 1 × 1 mm). For the control group, MRI images were acquired on a Siemens Magnetom Trio 3T at the Mind Research Network (Franco, Mannell, Calhoun, & Mayer, 2013; Zuo et al., 2014). These data were collected using a comparable scanning protocol consisting of a multi-echo MPRAGE sequence with acceleration factors on (IPAT = 2 GRAPPA) in interleaved order (192 slices, TR = 2530 ms, TE = 1.64, 3.5, 5.36, 7.22, 9.08 ms, FOV = 256 × 256 mm with voxel size = 1 × 1 × 1 mm).

2.6. SMRI data analyses

2.6.1. Longitudinal preprocessing

Structural MRI data were preprocessed using Freesurfer’s longitudinal processing stream (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999; Fischl, Sereno, Tootell, & Dale, 1999; Reuter et al., 2012). This processing stream includes resampling of each individual's MRI time points to a subject-specific baseline, which is a process that has been shown to increase both reliability and statistical power (Reuter et al., 2012). Further, the pipeline includes motion correction, skull-stripping, averaging T1 weighted images, removal of non-brain tissues, segmentation of gray matter, tessellation of white matter, automated topology correction, and parcellation according to gyral and sulcal structure. Freesurfer's calculations of CT allow for detection of submillimeter-level changes which are not constrained to the voxel size. All structural data were manually inspected and the CT maps were smoothed with full-width half maximum (FWHM) at 10 mm. Finally, Freesurfer's algorithms can accurately measure and model longitudinal data across different scanners and protocols, therefore decreasing the chances that between group differences for the L2 learners versus controls are due to scanning differences (Reuter et al., 2012).

2.6.2. ROI-based analyses

We used Freesurfer's univariate linear mixed-effects (LME) modeling (Bernal-Rusiel, Greve, Reuter, Fischl, & Sabuncu, 2013) to conduct a-priori designated ROI-based analyses of GMV and CT. This modeling toolkit allows for the examination of average cortical changes over time and allows for the examination of individual differences in behavioral performance and these cortical changes. The ROIs for the current study were derived from a previous fMRI connectivity study on these same L2 learning participants (Grant et al., 2015). We chose to examine whether cortical structural changes occurred in regions that gained or lost fMRI connectivity across L2 learning, as shown in Grant et al. (2015). Our CT ROIs included the bilateral ACC, IFG, MFG, and MTG, which were implicated in both the literature and the Grant et al. study as regions for cognitive and bilingual language control. GMV ROIs included the same regions as the CT ROIs with the addition of the CN.

MRI Session 1 (S1) versus Session 2 (S2) and Learning Group (L2 Learners vs Controls) were modeled as fixed effects, and subjects were entered as random effects. In order to control for differences in overall brain size across participants, estimated intracranial volume was entered as a variable of no interest in all our analyses. To examine relationships between individual differences in L2 learning performance and cognitive ability and cortical structure, we examined interactions between these behavioral performance measures and GMV and CT. For each model, Cook's distance values were computed to identify influential points. All analyses reported here were run with influential points removed. To correct for multiple comparisons, we used the Benjamin and Hochberg FDR correction (Benjamini & Hochberg, 1995). To estimate the power for group comparisons, we used the lme_realizedPower function available through Freesurfer's linear mixed effects modelling pipeline.

2.6.3. sMRI concordance with functional connectivity

Our previous study examined functional connectivity changes across L2 learning (Grant et al., 2015). This regression analysis sought to examine whether the degree of sMRI changes (as measured via percent increase in CT) in the left MTG and right ACC (the two regions that showed increase in CT over time) corresponded to the contemporaneous beta weights for functional connectivity across these two regions after L2 learning.

3. Results

3.1. Behavioral results

Data from the behavioral tasks are reported in Table 1. In general, we observe that participants' rate their L2 abilities relatively high (5 out of 7), although their adjusted mean accuracy on the TVIP is more in line with our characterization of the participants as intermediate learners. Although the LHQ and TVIP were only administered at the first session, participants completed the Semantic Judgment and Language Decision tasks at each session. For the Semantic Judgment task we see a speeding of responses at Session 2, as well as a reduction in variability as measured by the coefficient of variation (see Grant et al., 2015). For the Language Decision task, there is an interaction between Session and Word Type, such that Spanish words were identified the most quickly and accurately at both sessions, whereas English words were more accurately identified than Homographs at Session 1 but Homographs were more accurately identified than English words at Session 2. Although the increased accuracy and decreased RTs to the Spanish words compared with the English words may be counter-intuitive, this likely reflects a response bias caused by the greater percentage (67–33) of "Spanish" responses required by the single language decision task. In the context of the language decision task, participants were asked to only make 'yes' decisions to words they recognized as Spanish, and therefore responses to English words were rejections. It is a well-established finding that rejections are typically slower and less accurate (see Vuckovic, Kwantes, Humphreys, & Neal, 2014). The poorer performance for homograph words, as compared with unambiguous Spanish and English words, was expected given the conflict-inducing nature of the homographs (Van Heuven, Schriefers, Dijkstra, & Hagoort, 2008).

Table 1
Behavioral measures.

Measure	Accuracy % *M (SD)*	Reaction Time *M (SD)*
Proficiency		
TVIP	58% (7)	
*LHQ**	5 (0.65)	
Semantic judgment	S1: 82% (6)	S1: 926.1 ms (90.1)
	S2: 84% (4)	S2: 875.6 ms (105.8)
Language decision task		
Spanish Trials	S1: 97% (5)	S1: 740.1 ms (127.8)
	S2: 95% (6)	S2: 690.2 ms (59.3)
English Trials	S1: 91% (8)	S1: 786.3 ms (140.4)
	S2: 88% (8)	S2: 735.6 ms (78.5)
Homograph Trials	S1: 88% (6)	S1: 798.7 ms (186.9)
	S2:90% (9)	S2: 708.6 ms (67.0)
Cognitive Ability		
LNS	60% (14)	
*Flanker***		50.8 ms (21.3)

Note. S1 and S2 correspond to Session 1 and Session 2, respectively. *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.

3.2. sMRI group differences (Learners versus Controls)

Participants who were in the L2 Spanish classroom group showed greater CT in the right MTG and left ACC at Session 2, as compared to the control group who did not undergo any language learning between scans (Fig. 3; for the right MTG $F(1, 42) = 6.52, p = .01$, FDR corrected $p = .08$; for the left ACC $F(1, 42) = 5.82$, $p = .02$, FDR corrected $p = .08$). See Table 2 for all group level CT results. For GMV group level results, see Supplementary Table 1. There were no significant increases or decreases in CT or GMV in the control group (Table 3).

3.3. sMRI interactions with language performance

These analyses were conducted solely on the L2 Classroom learning participants because there were no language proficiency measures for

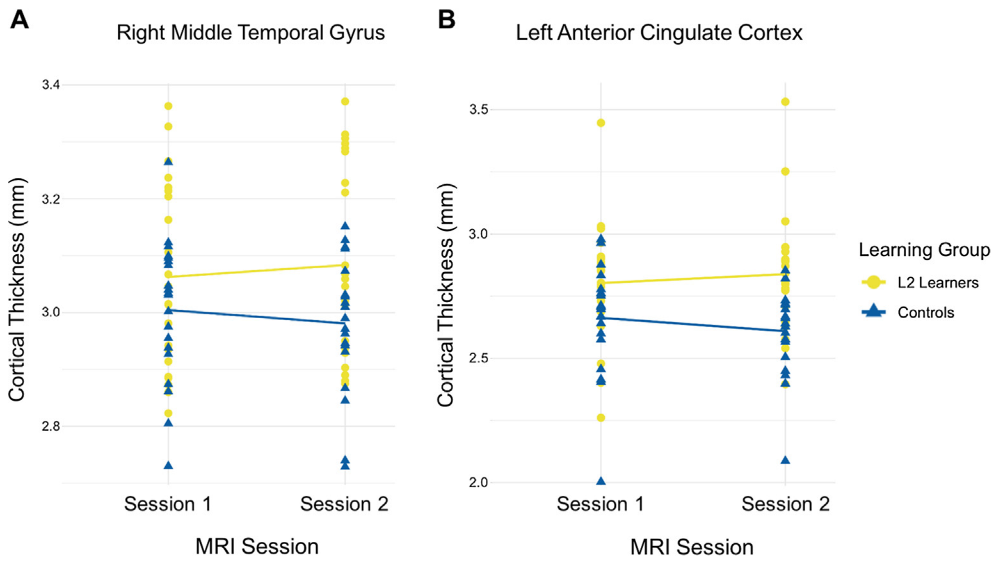

Fig. 3. At session 2, the L2 learning group showed greater CT in the (A) right MTG and (B) left ACC (FDR corrected $p = .08$).

Table 2
CT results for time × group interaction across L2 training group and controls.

Region of interest	F-value	Degrees of freedom	p-value	FDR adjusted p-value	Realized power
L ACC*	5.82	[1, 41.87]	0.02	0.08	0.65
R ACC	0.11	[1, 41.92]	0.74	0.84	0.06
L IFG	2.31	[1, 41.89]	0.14	0.36	0.32
R IFG	1.09	[1, 41.84]	0.30	0.60	0.18
L MFG	0.03	[1, 41.88]	0.85	0.85	0.05
R MFG	0.74	[1, 41.86]	0.39	0.63	0.13
L MTG	0.16	[1, 41.84]	0.69	0.84	0.07
R MTG*	6.52	[1, 41.79]	0.01	0.08	0.70

Note. Asterisks denote that CT increased more in the right MTG and left ACC for L2 training participants compared to controls after training, although these findings were marginal after FDR correction.

Table 3
Changes in CT for controls only.

Region of interest	F-value	Degrees of freedom	p-value	FDR adjusted p-value	Realized power
L ACC	3.35	[1, 18.05]	0.08	0.44	0.41
R ACC	1.63	[1, 18.02]	0.22	0.44	0.23
L IFG	0.47	[1, 18.03	0.50	0.67	0.10
R IFG	0.00	[1, 18.05]	0.96	0.96	0.05
L MFG	1.80	[1, 18.03]	0.20	0.44	0.24
R MFG	1.86	[1, 18.03]	0.19	0.44	0.25
L MTG	0.07	[1, 18.04]	0.79	0.90	0.06
R MTG	0.63	[1, 18.04]	0.44	0.67	0.12

Note. There were no significant increases or decreases in CT for the control groups over time.

the control group.

3.3.1. Spanish (L2) performance interactions

Participants with high L2 proficiency showed gray matter increase over time: (1) participants who showed high accuracy on the Spanish SSJ task showed increased GMV in the left IFG pars triangularis after L2 learning (Fig. 4; $F(1, 19) = 7.40$; $p = .015$; adjusted $p = .071$ after FDR correction) and (2) participants who learned their L2 earlier in life showed a significant increase in GMV in the right IFG pars triangularis (Fig. 5; $F(1, 20) = 8.13$; $p = .009$; adjusted $p = .001$ after FDR correction). To help visualize these changes across time, we plotted these relationships as the correlation between percent change in GMV with SSJ improvement scores and AoA in Figs. 4 and 5, respectively.

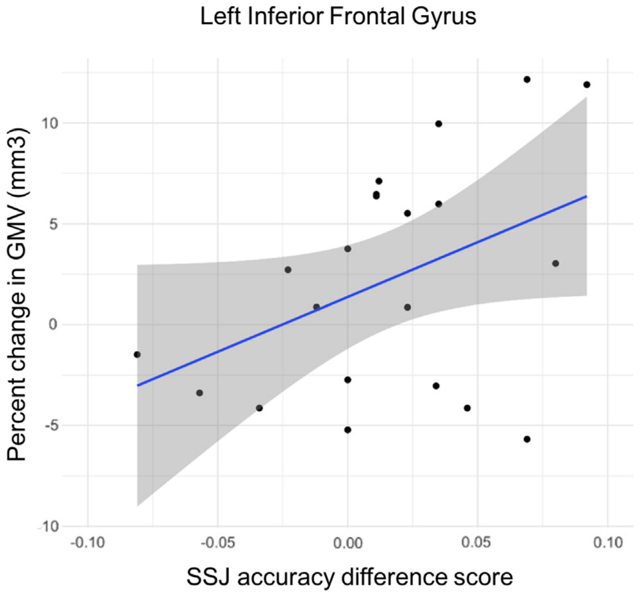

Fig. 4. Participants with high SSJ accuracy showed greater percent change in GMV in the left IFG across L2 learning (FDR corrected $p = .071$).

3.3.2. Homograph (L1 and L2) performance interactions

CT and GMV did not correlate significantly with homograph trials of the language decision task.

3.3.3. English discrimination performance interactions

Participants who were better able to distinguish English from Spanish words of the language decision (LD) task showed greater CT in the right MTG after L2 learning (Fig. 6; $F(1, 19) = 6.91$; $p = .014$; adjusted $p = .027$ after FDR correction). This was also plotted as the correlations between percent change in CT with LD accuracy scores.

3.4. sMRI interactions with cognitive ability

No measures of cognitive ability (i.e., neither the flanker nor the LNS task) were significantly correlated with CT or GMV changes over time.

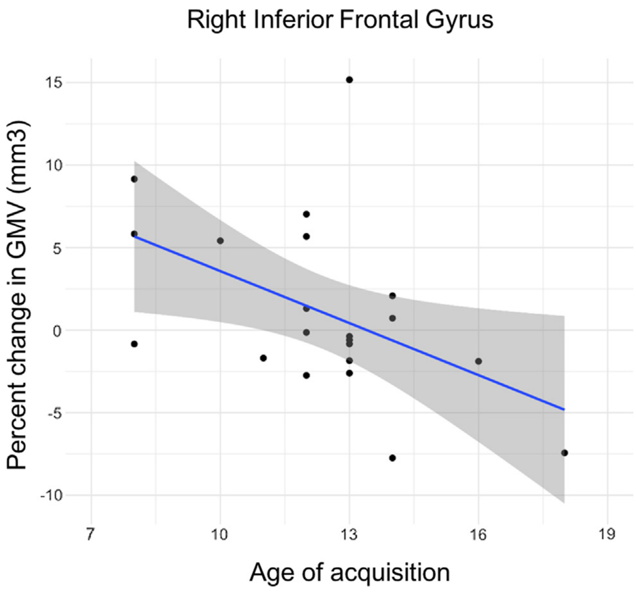

Fig. 5. Participants with an early age of acquisition showed greater percent change in GMV in the right IFG after L2 learning (FDR corrected $p < .05$).

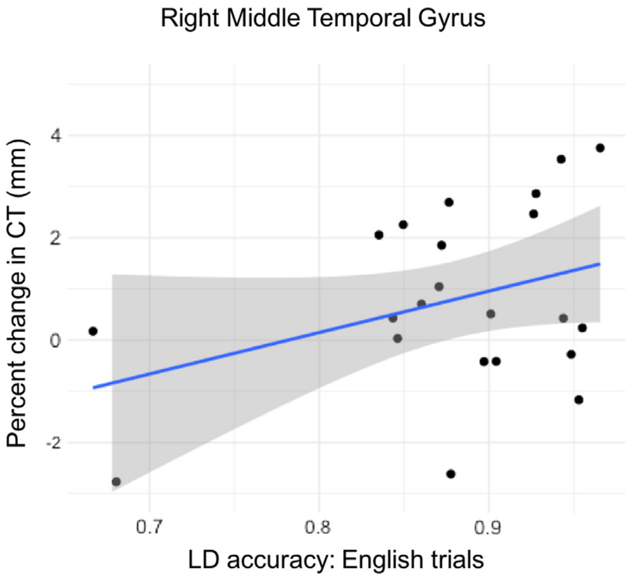

Fig. 6. Participants with high LD accuracy on distinguishing English (L1) from Spanish (L2) words showed greater CT in the right MTG after L2 learning (FDR corrected $p < .05$).

3.5. sMRI regressions with functional connectivity

In order to examine the relationship between sMRI and functional connectivity results from our previous study (Grant et al., 2015), we performed regressions for percent increase in the left ACC and right MTG and the beta weights between these regions. We found that percent increase in left ACC CT across L2 learning was positively correlated with contemporaneous connectivity beta weights between the ACC and MTG at session 2 ($F(1, 17) = 6.52$; $p = .021$; adjusted $p = .042$ after FDR correction). The percent increase in the right MTG was not significantly correlated with the functional connectivity beta weights between these two structures.

4. Discussion

The current study focuses on L1 and L2 lexical processing and its relationship with structural brain changes, specifically gray matter structure, as a function of Spanish classroom learning. In a previous fMRI connectivity study on the same L2 learning participants, participants showed changes in a functional network comprising the bilateral anterior cingulate cortex (ACC), caudate nucleus (CN), inferior frontal gyrus (IFG), middle frontal gyrus (MFG) and middle temporal gyrus (MTG; Grant et al., 2015). The main research goals of the current study were to examine longitudinal structural brain changes, as measured by CT and GMV, in these brain regions across L2 learning in order to gain a more systematic understanding of the relationships between L2 learning, structural and functional changes, and individual differences in language and cognitive control performance. Our findings support the hypotheses that (1) L2 learning leads to structural brain changes that correspond with a functional connectivity network and (2) these structural changes are correlated with individual differences in L1 and L2 performance and experience. To the best of our knowledge, this is the first study to examine whether CT changes correlate with task-based functional connectivity in healthy humans, indicating a possible relationship between these two measures.

4.1. L2 learning leads to increased CT in language and cognitive control regions

Our findings comparing L2 learners to controls are consistent with studies indicating CT and GMV increase with L2 learning (see Li et al., 2014 for a review). Specifically, L2 learners showed greater CT after learning as compared to controls in the right middle temporal gyrus (MTG) and left anterior cingulate cortex (ACC). Although these differences were marginal after FDR correction, we note that we used independent *a priori* designated regions of interest, which reduced the number of tests performed to eight. We include both the corrected and uncorrected values in the interests of full transparency, and suggest that given our targeted approach, the distinction between the uncorrected and corrected *p*-values is less critical than in a whole-brain analysis where thousands of tests are taking place (Lindquist & Mejia, 2015). By contrast, none of these regions increased or decreased significantly (either corrected or uncorrected) for the control group. Previous sMRI studies have found that the anterior temporal lobe, which contains portions of the MTG reported in this study, increased with L2 learning and is associated with increased L2 proficiency (Hosoda et al., 2013; Stein et al., 2012). The MTG has been implicated as an integral structure involved in lexical and semantic processing, semantic integration, and long-term storage of conceptual-semantic knowledge according to functional neuroimaging studies (Lambon Ralph, 2014; Rodríguez-Fornells et al., 2009), including L2 lexical processing in late L2 learners (Hernandez et al., 2015). GMV in the ACC has been shown to be associated with weaker behavioral conflict effect (associated with greater inhibitory control) in bilinguals as compared to monolinguals, indicating that conflict monitoring between L1 and L2 items may rely on the ACC (Abutalebi et al., 2012). Several functional MRI studies have shown a modulatory role of the ACC in cognitive control, selective attention, and language production (Abutalebi & Green, 2007). Taken together, these findings support the hypothesis that L2 learning leads to both structural and functional changes in brain regions implicated in L2 lexico-semantic and cognitive control. An interesting finding in itself was that we were only able to find differences in CT and not GMV for the L2 learners versus controls. This may be due to differences in how these measures account for (or fail to account for) levels of cortical folding. CT measures and GMV measures can often be at odds with one another in regions with high levels of cortical folding (Chung et al., 2006). Because the CT measure is able to account for cortical folding on a sub-millimeter level, it may be a more sensitive measure as compared to GMV at detecting between group differences.

We should note that we did not observe effects of language experience in the caudate nucleus. The lack of effects in this region was relatively unexpected, given (a) its prominence in current theories of language control, such as the Adaptive Control Hypothesis (Abutalebi & Green, 2007) and (b) the functional differences we observed in our previous study, wherein the connectivity between the caudate and the IFG as well as the ACC increased over time (Grant et al., 2015). Previous work suggests that the structure of the caudate expands at early stages of L2 acquisition and contracts at later stages or with extensive immersion experience (see Pliatsikas, DeLuca, Moschopoulou, & Saddy, 2017). Although one might expect, given the intermediate proficiency of our learners, to observe expansion in this region, there are several important differences between our learners and those previously studied. First, most of our learners (75%) had no immersion experience. Second, previous work directly compared bilinguals to monolinguals, whereas our longitudinal study makes a within-participants comparison. Consequently, there are two possible reasons we did not observe effects of language learning in the caudate. First, it is possible that we missed the opportunity to observe structural changes in that region by testing learners who were already at an intermediate stage in their acquisition. Future studies should, if possible, take a prospective approach and scan learners before any exposure to the L2. The second possibility is that L2 immersion experience plays a critical role in shaping the caudate, and consequently we did not observe effects due to the fact that our participants were classroom learners who were generally immersed in their L1.

4.2. Individual differences in L2 learning-associated neuroplasticity

The results of our study indicate that brain regions increased in CT and GMV only in participants who performed with high accuracy during L1 and L2 tasks. For example, participants who were better able to distinguish L1 from L2 (i.e., correctly reject English words when asked to identify Spanish words during the language decision task) showed increased CT in the right MTG after L2 learning. Further, GMV in the left IFG increased after L2 learning only for those who had high accuracy on the Spanish semantic judgement (SSJ) task (however, this result is marginal after multiple comparisons correction) and GMV in the right IFG was greater for those who learned their L2 at an earlier age versus later L2 learners. These findings are in line with the majority of structural imaging studies purporting a significant relationship between L2 experience and GMV in the IFG (Li et al., 2014), and consistent with functional studies associating the activity in the IFG with L2 learning, lexical retrieval, and bilingual language production (Parker Jones et al., 2012). Given our findings that CT increases in those who learned their L2 at an earlier AoA, this may suggest that participants who learn an L2 earlier in life may have a more flexible or plastic brain, structurally and functionally. In sum, both functional and structural findings indicate the IFG is sensitive to L2 proficiency and AoA (see also recent findings from Luo et al., 2019 and Nichols & Joanisse, 2016).

Our finding that participants' ability to distinguish L1 words from L2 words was positively related to their CT in the right MTG after L2 learning is consistent with previous studies that have implicated the MTG in both L1 and L2 semantic processing and comprehension (Marian, Spivey, & Hirsch, 2003; Perani et al., 1996; Rodríguez-Fornells et al., 2009). Since learning a second language often involves integrating some aspects of L1 and L2 vocabulary, it could be that participants who have greater ability to differentiate L1 from L2 words use the MTG to a higher degree when learning their L2. This would be consistent with the previous functional connectivity findings indicating enhanced reliance on the MTG, as shown by several additional connections to the MTG, after L2 learning (Grant et al., 2015). Together, these findings suggest that lexical development at the intermediate stage of L2 learning is associated with changes in brain structure and functional connectivity across regions implicated in semantic networks.

4.3. Relationships between structural and functional connectivity changes in L2 learning

Multimodal neuroimaging allows for a broader approach to understanding experience-dependent changes in the brain. In many cases of experience-based neuroplasticity, there is a strong structure-function correspondence. That is, experience-dependent changes in brain structure tend to occur in regions functionally implicated in the experience or task being learned (Green & Bavelier, 2008; Li et al., 2014; Lövdén et al., 2013). While the direction of this relationship has yet to be established – that is, whether the underlying structural changes lead to functional changes or vice versa, many studies have shown that there is a close relationship between gray matter structure as measured by either cortical thickness or gray matter volume and functional activity (see Li et al., 2014 for a review). However, this is the first language study to specifically examine the relationship between gray matter changes and functional connectivity.

Studies on rodents have shown a direct relationship between functional and structural reorganization in response to motor learning (Kleim et al., 2002), and graph theory approaches to CT have shown that CT networks are in line with resting-state functional connectivity networks in humans (Chen et al., 2008; He et al., 2007). Taking this into account, we suggest that task-based functional connectivity and cortical thickness may also be associated with one another. This would be in line with the Hebbian principle that "cells that fire together wire together", such that neurons that are functionally connected should also be structurally connected, and neurons that have more connections (synapses) should be captured by measures of CT and GMV, which are both aggregate measures of axonal architecture, cell body, and dendritic and synaptic changes (Zatorre et al., 2012). To support this perspective, we compared the results of the current sMRI study with the previous fMRI findings in the same participants (Grant et al., 2015). The majority of the regions shown to be functionally active and/or functionally connected in Grant et al. (2015) also underwent L2-learning associated changes in CT and GMV in the current study.

Specifically, our sMRI findings show a high degree of correspondence with the previous fMRI task-based functional connectivity results, which may suggest that regions that gain functional connections as a function of L2 experience also undergo structural changes, as follows: (1) the two main regions that gained functional connections after L2 learning in Grant et al.'s study also showed CT increases over time, with the caveat that these were marginal after FDR correction (2) the majority of our sMRI findings mirrored previous functional connectivity results and, in particular, (3) CT increase in the left ACC, a region implicated in cognitive control and language monitoring, was significantly positively associated with the beta weight of the functional connections between itself and the lexico-semantic hub of the MTG.

While we did not find any significant relationship between the percent increase in CT for the MTG and any individual beta weights, there are several possible explanations for this. For one, it is possible that the number of connections may play a role such that the relationship between CT and functional connectivity may be easier to capture when there are fewer new connections. Given the fact that the MTG gained three additional connections (with the IFG, medial frontal lobe, and ACC; see Grant et al., 2015 Fig. 4), it is possible that CT in the MTG is related to all three added connections and is not driven by one particular connection. In contrast, the ACC formed new connections with only two regions (the MTG and IFG). It is also possible (though speculative) that the direction of the functional connectivity may play a role such that regions that are predominantly acting as influencers (as in the case of the ACC, which formed new connections) may show more CT-related increases as compared to regions that act as receivers (as is the case with the MTG, which received new connections). These multimodal findings suggest that the integration of lexico-semantics and cognitive control may underlie successful intermediate-level lexical processing.

4.4. Interpreting structure in the context of the CH and BIA-d

As discussed in the Introduction, the BIA-d (Grainger et al., 2010) model predicts that successful semantic processing during later stages of L2 learning will be dependent on inhibition of L1-L2 word form connections in order to facilitate direct connections between L2 word forms and the semantic store. By contrast, the CH (Green, 2003) predicts decreasing dependency on cognitive control and re-focused resources on a language semantic network during proficient, late stage L2 learning. Previous analyses of the functional data from our L2 learning group found that adult L2 Spanish learners relied on language and cognitive control networks for the early stages of L2 learning and switched to relying more on semantic processing regions during later learning stages, in support of the CH. The structural analyses presented here provide further support for the CH in that we observe structural differences in the MTG, a semantic processing region, as well as the ACC, which is associated with conflict monitoring rather than inhibition (as would be predicted by the BIA-d). Further strengthening this position is the correspondence that we have observed between structural changes and functional connectivity, such that the increase in left ACC CT was significantly positively associated with the beta weight of the functional connections between itself and the MTG.

4.5. Conclusions

The current study examines CT and GMV changes in semantic processing and cognitive control networks across two semesters of Spanish classroom learning. This study presents the first longitudinal multimodal approach towards examining the neurodevelopmental aspects of lexical processing across late L2 learners, and emphasizes the behavioral factors that might be associated with these brain changes and L2 experience. In general, our findings suggest that L2 learning-associated may lead to increases in CT in the right MTG and left ACC, two key brain regions involved in lexico-semantics and cognitive control that have been implicated in both previous literature and in our own fMRI study (Grant et al., 2015). Further, our sMRI findings corroborate previous fMRI studies emphasizing the role of L2 experience in GMV changes in the IFG and MTG. Moreover, our findings suggest a novel relationship between CT in the ACC and task-based functional connectivity changes between the ACC and MTG in response to intermediate level L2 learning. Altogether, these findings provide considerable support to the idea that effective L2 learners' lexical development relies on the collaboration of cognitive control and semantic networks.

Funding

This research reported in this study was supported by a research grant from the National Science Foundation (NSF-FO# 1533625) to PL.

Acknowledgements

We would like to thank Ying-An Chi and Weitao Chen for their assistance in data coding. In addition, we would like to thank the Penn State Social, Life, and Engineering Sciences Imaging Center MRI facility for the use of in-kind hours during data collection.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2019.104661.

References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica, 128*(3), 466–478. https://doi.org/10.1016/j.actpsy.2008.03.014.

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., ... Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex, 22*(9), 2076–2086. https://doi.org/10.1093/cercor/bhr287.

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*(3), 242–275. https://doi.org/10.1016/j.jneuroling.2006.10.003.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A pratical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289–300.

Berken, J. A., Gracco, V. L., Chen, J. K., & Klein, D. (2016). The timing of language learning shapes brain structure associated with articulation. *Brain Structure and Function, 221*(7), 3591–3600.

Bernal-Rusiel, J. L., Greve, D. N., Reuter, M., Fischl, B., & Sabuncu, M. R. (2013). Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. *NeuroImage, 66*, 249–260. https://doi.org/10.1016/j.neuroimage.2012.10.065.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001.

Chai, X. J., Berken, J. A., Barbeau, E. B., Soles, J., Callahan, M., Chen, J. K., & Klein, D. (2016). Intrinsic functional connectivity in the adult brain and success in second-language learning. *Journal of Neuroscience, 36*(3), 755–761.

Chen, Z. J., He, Y., Rosa-Neto, P., Germann, J., & Evans, A. C. (2008). Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cerebral Cortex, 18*(10), 2374–2381. https://doi.org/10.1093/cercor/bhn003.

Chung, M. K., Dalton, K. M., Shen, L., Evans, A. C., & Davidson, R. J. (2006). Unified cortical surface morphometry and its application to quantifying amount of gray matter. Department of Statistics, U. of Wisonsin-Madison, (Techinical Report No. 1122).

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage, 9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395.

Della Rosa, P. A., Videsott, G., Borsa, V. M., Canini, M., Weekes, B. S., Franceschini, R., & Abutalebi, J. (2013). A neural interactive location for multilingual talent. *Cortex, 49*(2), 605–608. https://doi.org/10.1016/j.cortex.2012.12.001.

DeLuca, V., Rothman, J., & Pliatsikas, C. (2018). Linguistic immersion and structural effects on the bilingual brain: A longitudinal study. *Bilingualism: Language and Cognition*, 1–16.

Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *Test de Vocabulario en Imágenes Peabody – Adaptación Hispanoamericano.* Circle Pines, MN: American Guidance Service.

Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science, 19*(12), 1201–1206.

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences, 97*(20), 11050–11055. https://doi.org/10.1073/pnas.200033797.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. *NeuroImage, 9*(2), 195–207. https://doi.org/10.1006/nimg.1998.0396.

Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-resolution intersubject averaging and a surface-based coordinate system. *Human Brain Mapping, 8*(February 1999), 272–284. https://doi.org/10.1002/(SICI)1097-0193(1999) 8.

Franco, A. R., Mannell, M. V., Calhoun, V. D., & Mayer, A. R. (2013). Impact of analysis methods on the reproducibility and reliability of resting-state networks. *Brain Connectivity, 3*(4), 363–374. https://doi.org/10.1089/brain.2012.0134.

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail, & M. Hickmann (Eds.). *Language acquisition across linguistic and cognitive systems* (pp. 267–284). New York, NY: John Benjamins.

Grant, A., Fang, S. Y., & Li, P. (2015). Second language lexical development and cognitive control: A longitudinal fMRI study. *Brain and Language, 144*, 35–47. https://doi.org/10.1016/j.bandl.2015.03.010.

Green, D. W. (2003). Neural basis of lexicon and grammar in L2 acquisition: The convergence hypothesis. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.). *The interface between syntax and the lexicon in second language acquisition* (pp. 197–208). Amsterdam: John Benjamins.

Green, C. S., & Bavelier, D. (2008). Exercising your brain: A review of human brain plasticity and training-induced learning. *Psychology and Aging, 23*(4), 692–701. https://doi.org/10.1037/a0014345.

Green, D. W., Crinion, J., & Price, C. J. (2006). Convergence, degeneracy and control. Retrieved from *Language Learning, 56*, 1–16. http://doi.wiley.com/10.1111/j.1467-9922.2006.00357.x%5Cnpapers3://publication/doi/10.1111/j.1467-9922.2006.00357.x.

Grogan, A., Green, D. W., Ali, N., Crinion, J. T., & Price, C. J. (2009). Structural correlates of semantic and phonemic fluency ability in first and second languages. *Cerebral Cortex, 19*(11), 2690–2698. https://doi.org/10.1093/cercor/bhp023.

Grogan, A. O., Parker Jones, T., Ali, N., Crinion, J., Orabona, S., Mechias, M. L., ... Price, C. J. (2012). Structural correlates for lexical efficiency and number of languages in non-native speakers of English. *Neuropsychologia, 50*(7), 1347–1352. https://doi.org/10.1016/j.neuropsychologia.2012.02.019.

He, Y., Chen, Z. J., & Evans, A. C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cerebral Cortex, 17*(10), 2407–2419. https://doi.org/10.1093/cercor/bhl149.

Hernandez, A. E. (2013). *The bilingual brain.* Oxford University Press.

Hernandez, A., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133*(4), 638–650. https://doi.org/10.1037/0033-

2909.133.4.638.

Hernandez, A. E., Woods, E. A., & Bradley, K. A. L. (2015). Neural correlates of single word reading in bilingual children and adults. *Brain and Language, 143*, 11–19. https://doi.org/10.1016/j.bandl.2015.01.010.

Hervais-Adelman, A., Moser-Mercer, B., Murray, M. M., & Golestani, N. (2017). Cortical thickness increases after simultaneous interpretation training. *Neuropsychologia, 98*, 212–219.

Hofstetter, S., Friedmann, N., & Assaf, Y. (2017). Rapid language-related plasticity: Microstructural changes in the cortex after a short session of new word learning. *Brain Structure and Function, 222*(3), 1231–1241.

Hosoda, C., Tanaka, K., Nariai, T., Honda, M., & Hanakawa, T. (2013). Dynamic neural network reorganization associated with second language vocabulary acquisition: A multimodal imaging study. *Journal of Neuroscience, 33*(34), 13663–13672. https://doi.org/10.1523/JNEUROSCI.0410-13.2013.

Kleim, J. A., Barbay, S., Cooper, N. R., Hogg, T. M., Reidel, C. N., Remple, M. S., & Nudo, R. J. (2002). Motor learning-dependent synaptogenesis is localized to functionally reorganized motor cortex. *Neurobiology of Learning and Memory, 77*(1), 63–77. https://doi.org/10.1006/nlme.2000.4004.

Klein, D., Mok, K., Chen, J. K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language, 131*, 20–24.

Lambon Ralph, M. A. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. 20120392–20120392 *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), https://doi.org/10.1098/rstb.2012.0392.

Lemaitre, H., Goldman, A. L., Sambataro, F., Verchinski, B. A., Meyer-Lindenberg, A., Weinberger, D. R., & Mattay, V. S. (2012). Normal age-related brain morphometric changes. *Neurobiology of Aging, 33*(3), 1–15. https://doi.org/10.1016/j.neurobiolaging.2010.07.013.Normal.

Li, P., & Grant, A. (2015). Identifying the causal link: Two approaches toward understanding the relationship between bilingualism and cognitive control. *Cortex, 73*, 358–360.

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex, 58*, 301–324. https://doi.org/10.1016/j.cortex.2014.05.001.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A Web based interface for bilingual research. *Behavior Research Methods, 38*(2), 202–210.

Lindquist, M. A., & Mejia, A. (2015). Zen and the art of multiple comparisons. *Psychosomatic Medicine, 77*(2), 114.

Lövdén, M., Wenger, E., Mårtensson, J., Lindenberger, U., & Bäckman, L. (2013). Structural brain plasticity in adult learning and development. *Neuroscience and Biobehavioral Reviews, 37*(9), 2296–2310. https://doi.org/10.1016/j.neubiorev.2013.02.014.

Luo, D., Kwok, V. P., Li, P., Liu, Q., Li, W., Yang, Y., ... Tan, L. H. (2019). Microstructural plasticity in the bilingual brain. *Brain and Language, 196*, 1–9.

Mamiya, P. C., Richards, T. L., Coe, B. P., Eichler, E. E., & Kuhl, P. K. (2016). Brain white matter structure and COMT gene are linked to second-language learning in adults. *Proceedings of the National Academy of Sciences, 113*(26), 7249–7254.

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language, 86*(1), 70–82. https://doi.org/10.1016/S0093-934X(02)00535-7.

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage, 63*(1), 240–244. https://doi.org/10.1016/j.neuroimage.2012.06.043.

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Neurolinguistics: Structural plasticity in the bilingual brain. *Nature, 431*(7010), 757. https://doi.org/10.1038/431757a.

Mollink, J., Smith, S. M., Elliott, L. T., Kleinnijenhuis, M., Hiemstra, M., Alfaro-Almagro, F., ... Miller, K. L. (2019). The spatial correspondence and genetic influence of interhemispheric connectivity with white matter microstructure. *Nature Neuroscience, 1*.

Narr, K. L., Bilder, R. M., Toga, A. W., Woods, R. P., Rex, D. E., Szeszko, P. R., ... Thompson, P. M. (2005). Mapping cortical thickness and gray matter concentration in first episode schizophrenia. *Cerebral Cortex, 15*(6), 708–719. https://doi.org/10.1093/cercor/bhh172.

Narr, K. L., Woods, R. P., Thompson, P. M., Szeszko, P., Robinson, D., Dimtcheva, T., ... Bilder, R. M. (2007). Relationships between IQ and regional cortical gray matter thickness in healthy adults. *Cerebral Cortex, 17*(9), 2163–2171. https://doi.org/10.1093/cercor/bhl125.

Nichols, E. S., & Joanisse, M. F. (2016). Functional activity and white matter microstructure reveal the independent effects of age of acquisition and proficiency on secondlanguage learning. *Neuroimage, 143*, 15–25.

Parker Jones, O., Green, D. W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., ... Price, C. J. (2012). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex, 22*(4), 892–902. https://doi.org/10.1093/cercor/bhr161.

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., ... Mehler, J. (1996). Brain processing of native and foreign languages. *NeuroReport, 7*(15), 2439–2444.

Pliatsikas, C. (2019). Understanding structural plasticity in the bilingual brain: The dynamic restructuring model. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S1366728919000130 (in press).

Pliatsikas, C., DeLuca, V., Moschopoulou, E., & Saddy, J. D. (2017). Immersive bilingualism reshapes the core of the brain. *Brain Structure and Function, 222*(4), 1785–1795.

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences, 1191*, 62–88. https://doi.org/10.1111/j.1749-6632.2010.05444.x.

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage, 61*(4), 1402–1418. https://doi.org/10.1016/j.neuroimage.2012.02.084.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364*(1536), 3711–3735.

Rosas, H. D., Liu, A. K., Hersch, S., Glessner, M., Ferrante, R. J., Salat, D. H., ... Fischl, B. (2002). Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology, 58*(5), 695–701. https://doi.org/10.1212/WNL.58.5.695.

Sheppard, J. P., Wang, J.-P., & Wong, P. C. M. (2012). Large-scale cortical network properties predict future sound-to-word learning success. *Journal of Cognitive Neuroscience, 24*(5), 1087–1103. https://doi.org/10.1162/jocn_a_00210.

Stein, M., Federspiel, A., Koenig, T., Wirth, M., Strik, W., Wiest, R., ... Dierks, T. (2012). Structural plasticity in the language system related to increased second language proficiency. *Cortex, 48*(4), 458–465. https://doi.org/10.1016/j.cortex.2010.10.007.

Stein, M., Winkler, C., Kaiser, A. C., & Dierks, T. (2014). Structural brain changes related to bilingualism: Does immersion make a difference? *Frontiers in Psychology, 5*(Sep), 1–7. https://doi.org/10.3389/fpsyg.2014.01116.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, 94*(26), 14792–14797.

Van de Putte, E., De Baene, W., García-Pentón, L., Woumans, E., Dijkgraaf, A., & Duyck, W. (2018). Anatomical and functional changes in the brain after simultaneous interpreting training: A longitudinal study. *Cortex, 99*, 243–257.

Van Heuven, W. J., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex, 18*(11), 2706–2716.

Vuckovic, A., Kwantes, P. J., Humphreys, M., & Neal, A. (2014). A sequential sampling account of response bias and speed–accuracy tradeoffs in a conflict detection task. *Journal of Experimental Psychology: Applied, 20*(1), 55–68.

Wechsler, D. (1997). *Wechsler adult intelligence scale-(WAIS-III)*. San Antonio: Psychological Corporation.

Xiang, H., van Leeuwen, T. M., Dediu, D., Roberts, L., Norris, D. G., & Hagoort, P. (2015). L2-proficiency-dependent laterality shift in structural connectivity of brain language pathways. *Brain Connectivity, 5*(6), 349–361.

Yang, J., Gates, K. M., Molenaar, P., & Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics, 33*, 29–49. https://doi.org/10.1016/j.jneuroling.2014.09.004.

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: Neuroimaging changes in brain structure during learning. *Nature Neuroscience, 15*(4), 528–536. https://doi.org/10.1038/nn.3045.

Zou, L., Ding, G., Abutalebi, J., Shu, H., & Peng, D. (2012). Structural plasticity of the left caudate in bimodal bilinguals. *Cortex, 48*(9), 1197–1206. https://doi.org/10.1016/j.cortex.2011.05.022.

Zuo, X. N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., ... Milham, M. P. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific Data, 1*, 1–13. https://doi.org/10.1038/sdata.2014.49.