



Foreign Language Learning in Older Adults: Anatomical and Cognitive Markers of Vocabulary Learning Success

Manson Cheuk-Man Fong^{1,2*}, Matthew King-Hang Ma³, Jeremy Yin To Chui¹, Tammy Sheung Ting Law¹, Nga-Yan Hui¹, Alma Au⁴ and William Shiyuan Wang^{1,2,3*}

¹ Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China, ² Research Institute for Smart Ageing, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China, ³ Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China, ⁴ Department of Applied Social Science, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China

OPEN ACCESS

Edited by:

Toshiharu Nakai,
Osaka University, Japan

Reviewed by:

Giovanna Bubbico,
University of Studies G. d'Annunzio
Chieti and Pescara, Italy
Olga Kepinska,
University of Vienna, Austria
Eri Nakagawa,
National Institute for Physiological
Sciences (NIPS), Japan

*Correspondence:

Manson Cheuk-Man Fong
cmmfong@polyu.edu.hk
William Shiyuan Wang
wsywang@polyu.edu.hk

Specialty section:

This article was submitted to
Speech and Language,
a section of the journal
Frontiers in Human Neuroscience

Received: 30 September 2021

Accepted: 08 February 2022

Published: 07 March 2022

Citation:

Fong MC-M, Ma MK-H, Chui JYT,
Law TST, Hui N-Y, Au A and
Wang WS (2022) Foreign Language
Learning in Older Adults: Anatomical
and Cognitive Markers of Vocabulary
Learning Success.
Front. Hum. Neurosci. 16:787413.
doi: 10.3389/fnhum.2022.787413

In recent years, foreign language learning (FLL) has been proposed as a possible cognitive intervention for older adults. However, the brain network and cognitive functions underlying FLL has remained largely unconfirmed in older adults. In particular, older and younger adults have markedly different cognitive profile—while older adults tend to exhibit decline in most cognitive domains, their semantic memory usually remains intact. As such, older adults may engage the semantic functions to a larger extent than the other cognitive functions traditionally considered the most important (e.g., working memory capacity and phonological awareness). Using anatomical measurements and a cognitive test battery, the present study examined this hypothesis in twenty cognitively normal older adults (58–69 years old), who participated in a two-month Italian learning programme. Results showed that the immediate learning success and long-term retention of Italian vocabularies were most consistently predicted by the anatomical measures of the left pars orbitalis and left caudal middle frontal cortex, which are implicated in semantic and episodic memory functions. Convergent evidence was also found based on the pattern of cognitive associations. Our results are consistent with a prominent role of semantic and episodic memory functions in vocabulary learning in older learners.

Keywords: foreign language learning, vocabulary learning, structural MRI, FreeSurfer, pars orbitalis, caudal middle frontal cortex, semantic memory, episodic memory

1. INTRODUCTION

1.1. Foreign Language Learning in Older Adults

In foreign language learning (FLL) research, older adults have been an understudied population (Mackey and Sachs, 2012), which may be caused by the increased difficulty in picking up a new language after the “critical period” (Lenneberg, 1967; Hartshorne et al., 2018, see also Wang, 2018). However, there has been a recent change in attitude, due in part to a series of pioneering studies on the effect of lifelong bilingualism on brain structures and cognitive functions (Bialystok et al., 2007; Luk et al., 2011; Costa and Sebastián-Gallés, 2014; Olsen et al., 2015) and on the neuroplasticity induced by intensive FLL in younger adults (Wong and Perrachione, 2007; Mårtensson et al., 2012; Stein et al., 2012; Zatorre, 2013; Qi et al., 2019).

Relative to older monolinguals, older bilinguals were reported to have greater protection against dementia, better inhibition function, and improved executive control (Bialystok et al., 2007; van den Noort et al., 2019). The bilingual advantage has also been reported for other cognitive domains, e.g., episodic and verbal memory (Schroeder and Marian, 2012; Grant et al., 2014, but see Paap et al., 2015). Neurally, the correlates of bilingual experiences have been studied using many anatomical segmentation tools (Li et al., 2014; Hämäläinen et al., 2018; Maschio et al., 2019). Using a technique known as surface-based morphometry (SBM; Fischl and Dale, 2000; Winkler et al., 2010; Luders et al., 2012), early acquisition of two languages is found to be associated with larger surface area in the left pars opercularis of the inferior frontal gyrus (IFG) and right superior temporal gyrus (STG), while late acquisition is associated with increased mean curvature in the left STG (Hämäläinen et al., 2018). Bilingualism-induced neural differences have also been demonstrated using related anatomical measures such as sub-cortical volumes (DeLuca et al., 2019b) or gray matter density (Abutalebi et al., 2015), obtained by volume-based morphometry and voxel-based morphometry. For example, an expansion of the thalamus was observed for individuals with greater immersion in the second language (Pliatsikas et al., 2017; Deluca et al., 2019a).

As for learning-induced neuroplasticity in younger adults, five months of intensive FLL was sufficient to bring about increases in gray matter volume in the left IFG and left anterior temporal lobe (Stein et al., 2012). Intensive vocabulary learning in young interpreters was found to promote increases in cortical thickness across various regions in the left-lateralized language network, including left middle frontal gyrus, IFG, and STG, along with increased volume in the bilateral hippocampus (Mårtensson et al., 2012). Consistent with the classroom-based learning above, a recent experimental work based on either the picture–word association or virtual reality as the FLL paradigm revealed both structural and functional changes in younger adults, including increased cortical thickness over both the core language areas in the left hemisphere and many regions in the right hemisphere (Legault et al., 2019a). In another study, increased brain volume was found in the right hippocampus (Bellander et al., 2016).

Taken together, the findings on the bilingual advantage and learning-induced neuroplasticity have raised the possibility to use FLL as a non-pharmacological means for promoting successful aging in older adults (Antoniou et al., 2013; Antoniou and Wright, 2017; Ware et al., 2017; Klimova, 2018; Pfenninger and Polz, 2018; Klimova and Pikhart, 2020). Consistent with this conjecture, a few studies reported cognitive improvements in older learners after intensive FLL (Bak et al., 2016; Bubbico et al., 2019; Wong et al., 2019). For example, improvement in global cognition was found following six months of individual computer-based cognitively stimulating activities, including FLL and cognitive games (Wong et al., 2019). In another study, older learners also showed significant improvement in global cognition, along with increased resting-state functional connectivities (RSFCs) between posterior cingulate cortex and three right-hemispheric regions, including right IFG, right superior frontal gyrus, and left superior parietal lobule (Bubbico et al., 2019). This interesting finding suggests that

the right-hemisphere homologs of the left-hemisphere language areas may provide “scaffolding” in the early phase of second language learning, before a right-to-left shift occurs (Yang et al., 2015; Qi and Legault, 2020). Despite the positive findings above, results have been mixed, with a few studies reporting no improvement for all cognitive skills examined (Ware et al., 2017; Berggren et al., 2020), some degree of enhancement that was however not significantly different from that of a control group (Valis et al., 2019), or neither learning-induced anatomical nor behavioral changes (Nilsson et al., 2021).

1.2. Neurocognitive Bases of Foreign Language Learning

To account for the neuroplastic changes or cognitive gains that could be driven by FLL in older adults, an understanding about the neurocognitive bases of FLL is essential, because the learning-induced changes in a brain region likely arise due to its increased recruitment over the learning programme (Wong et al., 2007). Toward this end, some prominent neuroanatomical models of language (e.g., Ullman, 2001a, 2004; Hickok and Poeppel, 2007) have provided the guiding theoretical framework. In the declarative–procedural model (DP model; Ullman, 2001a, 2004), the declarative memory (comprising semantic and episodic memory), as subserved by a cortical–hippocampal system (McClelland et al., 1995; Eichenbaum, 2000), is posited to be involved for learning words. In contrast, the procedural memory, as subserved by the insula and basal ganglia (especially the striatum), is recruited in phonological and rule-based grammar learning. Functional activation and connectivities studies have been central to the quest for the neurocognitive bases of FLL (e.g., Mestres-Missé et al., 2008; Abutalebi and Green, 2016; Bakker-Marshall et al., 2021, for a recent review, see Qi and Legault, 2020). However, one source of information that could not be understated is individual differences—the focus of the present article. Traditionally, they are sometimes considered as “noise” that may obscure the effect of interest. Nonetheless, they have proved reliable for linking brain functions and cognitive behavior (e.g., Kanai and Rees, 2011; Chiarello et al., 2013).

In FLL research, individual differences in brain functions are posited to put strong constraints on the learning success (Zatorre, 2013; Li and Grant, 2016). Previous behavioral and neural findings have suggested that phonological skills, such as phonological short-term memory (PSTM) and phonological awareness, are the determining factors of language learning success (Papagno et al., 1991; Baddeley et al., 1998; Koda, 1998; Linck et al., 2014; Gillon, 2017). Corroborating evidence has been found in both classroom-based and laboratory-based studies with younger learners, for phonemic learning (Golestani and Zatorre, 2009; Silbert et al., 2015; Fuhrmeister and Myers, 2021), lexical tone learning (Wong and Perrachione, 2007; Chandrasekaran et al., 2010), and artificial grammar learning (Yang and Li, 2012; Kepinska et al., 2017). In agreement with the view that PSTM plays a crucial role in FLL, functional connectivity or activation studies have reported an association between FLL success and regions important for phonological processing, e.g., left precentral gyrus (Veroude et al., 2010), left

supramarginal gyrus (SMG; Veroude et al., 2010; Sliwiska et al., 2012), and left posterior STG (e.g., Wong et al., 2007). Also, consistent with the hypothesis that phonological awareness plays a critical role in language learning success, individual variation in learning outcomes can be predicted by individual variation in baseline sensitivity to non-lexical pitch patterns (Wong et al., 2007).

Of special interest to the present study are the anatomical associations with learning performance. For example, the success in linguistic pitch learning or direction discrimination were predicted by the volume of left Mescal's gyrus (Wong et al., 2008) and by the cortical thickness of the right hemispheric homolog of the Broca's area (Novén et al., 2019). In the study by Legault et al. (2019a), reviewed in the last section, the FLL performance was associated with the cortical thickness of right IFG in the picture-word group, and with the cortical thickness of right inferior parietal lobe in the VR group. The emphasis of phonological skills has extended to sub-cortical regions. Recent empirical works (Legault et al., 2019a) and theoretical papers (Abutalebi and Green, 2007; Li et al., 2014) have highlighted the roles of the striatum—comprising caudate nucleus and putamen—in phonological learning. For example, the caudate nucleus is involved in phonemic fluency (rather than semantic fluency; Grogan et al., 2009) and procedural memory (Ullman, 2004), while the putamen has been implicated in articulatory planning and detecting phonological errors (e.g., Abutalebi and Green, 2007).

While the emphasis on phonological skills in previous studies is understandable, in light of the DP model, it is surprising that the roles played by semantic and verbal episodic memory have seldom been tested in neural research on FLL. Since the DP model was initially proposed, functional imaging and neurological studies have established that semantic functions are supported by a widespread network that includes the ventrolateral prefrontal cortex, anterior temporal lobe, middle temporal gyrus, etc. (Patterson et al., 2007; Binder et al., 2009; Ralph et al., 2017), along with sub-cortical regions like the thalamus. In addition to being a main relay station of sensory information, the thalamus may also play an important role in language production by selecting lexical and semantic representations (Abutalebi and Green, 2016), by virtue of its connection with the left IFG (Ford et al., 2013). The episodic memory is subserved by the hippocampus and its surrounding medial temporal lobe structures (including the entorhinal cortex and the parahippocampal gyrus), along with the prefrontal cortex (e.g., Breitenstein et al., 2005). The hippocampus is well-known for its role in transforming short-term memory to long-term memory, while the entorhinal cortex and the parahippocampal gyrus are implicated during the processing of object/event information and spatial-temporal context, respectively (Eichenbaum et al., 2012). Another study also reported that the entorhinal cortex reflects elementary memory processes related to novelty detection, while the parahippocampus is more involved in the formation and subsequent reactivation of the memory (Daselaar et al., 2004). A few recent studies have reported evidence that supports a role of hippocampus in younger adults in FLL (Kepinska et al., 2017, 2018). For example, the connectivity of the hippocampus and

Broca's area was implicated during the acquisition of a novel grammar (Kepinska et al., 2018).

There are reasons that the brain networks and cognitive mechanisms could be different in older learners, which could result in different learning outcomes (Service and Craik, 1993; Glass et al., 2012; Ingvalson et al., 2017). To start with, older adults had poorer speech discrimination for phonemes and pitch (Shen et al., 2016), which may promote the use of alternative strategies in vocabulary learning. Due to lower working memory capacity (Service and Craik, 1993; Mackey and Sachs, 2012), they may also rely less on rote memorization. In addition, the inhibition function is generally in decline (Hasher et al., 1991, but see Verissimo et al., 2021), making it difficult for them to retrieve the L2 word because it is more difficult for them to suppress the L1 word. In contrast, older adults often have better crystallized knowledge (e.g., world and vocabulary knowledge) than younger adults (Salthouse, 2012; Hartshorne and Germine, 2015). It is plausible that declarative memory may play a crucial role in older learners (Ullman, 2001b). For example, they may take advantage of their better semantic knowledge to remember the vocabularies as a chunk (e.g., by forming a story to connect the words being learnt), or to proactively derive semantic associations for use as retrieval cues in episodic memory recall. However, individual differences studies on FLL in older adults have remained scarce, and even fewer have examined the neural correlates of FLL in older adults. To our knowledge, one such study has focused on vocabulary learning, and it suggested that the hippocampal volume and the associative memory ability prior to language learning are robust predictors of vocabulary proficiency at the end of training (Nilsson et al., 2021). Another study showed that the artificial grammar learning performance was influenced by structural and functional connectivity of the Broca's area and its right hemisphere homolog (Antonenko et al., 2012).

1.3. The Present Study

The present study investigates the neurocognitive factors that influence the immediate learning success and long-term retention of the vocabularies during an intensive FLL programme, based on a group of Cantonese-speaking Hong Kong older learners aged 58–69, all of whom were familiar with English as a second language. Italian was chosen as the target language due to its highly regular grapheme–phoneme correspondence, its use of the Latin alphabet, the unfamiliarity with the language by the general population in Hong Kong, and the popularity of Italy as a tourist destination. By design, while both vocabulary and basic grammar rules were taught throughout the learning programme, a stronger emphasis was put on the vocabulary since it plays a fundamental role in both spoken and written language comprehension across all stages of learning (Yum et al., 2014). Also, older learners may be less motivated in learning a new language in its entirety, and they may only wish to learn the essential vocabularies for simple communications and travels (Antoniou et al., 2013; Pfenninger and Polz, 2018). Our main hypothesis is that, for foreign vocabulary learning in older adults, both semantic and episodic memory are more strongly associated with the learning success than phonological skills. To examine this hypothesis, two groups

of analyses were conducted, using MRI and cognitive data to predict the immediate learning success and long-term retention.

The first group of analyses (MRI analysis) comprised a *cortical analysis* and a complementary *sub-cortical analysis*. For the cortical analysis, using SBM, four cortical morphological measures (cortical thickness, cortical surface area, cortical volume, and mean curvature) were extracted over a broad range of regions of interest (ROIs); these measures were then used as predictors to test the associations of each ROI with the in-class performances and the final test scores. The ROIs were selected because they have consistently been found to support semantic memory (Patterson et al., 2007; Lau et al., 2008; Binder et al., 2009; Ralph et al., 2017), have exhibited learning-induced neuroplasticity (Lee et al., 2007; Stein et al., 2012; Li et al., 2014), have previously been associated with second language learning performance in general (Ullman, 2016; Tagarelli et al., 2019), or are part of the language control network (Abutalebi and Green, 2007). Although language processing is strongly left-lateralized in the brain, the right hemisphere is also heavily engaged for learning a second language (Hosoda et al., 2013; Bubbico et al., 2019; Qi et al., 2019; Chen et al., 2021). Thus, the homologous areas in both hemispheres were included as ROIs. Our hypothesis would be supported if the cortical measures of the brain regions centrally involved in semantic functions (pars orbitalis and temporal pole) or episodic memory (caudal middle frontal cortex and entorhinal cortex) are more associated with learning performance than those that underlie phonological functions (precentral gyrus and SMG).

For the *sub-cortical analysis*, sub-cortical volume measures were used to predict in-class and final test scores. Due to our relatively small sample size and the already extended set of cortical regions examined, exploration was limited to four ROIs: hippocampus, thalamus, caudate nucleus, and putamen. This analysis complemented the cortical analysis in providing further evidence consistent with the main hypothesis. For example, if the hippocampal volume was associated with learning performance, it would strengthen the view that the episodic memory is associated with FLL in older adults.

In the second group of analyses (*cognitive analysis*), a broad range of cognitive and phonological measures, derived from a cognitive and phonological test battery, were used as the predictors of language learning performance. Behavioral support for our main hypothesis would be obtained if the semantic function score (semantic fluency and picture naming) and verbal episodic memory score (Hong Kong List Learning test) are more important predictors of the learning outcomes than phonological function scores (phonological discrimination and awareness).

2. MATERIALS AND METHODS

2.1. Participants

A group of 25 older learners aged 58–69 were initially enrolled into the Italian learning programme. They were native Cantonese speakers, had no known neurological disorders and normal/correct-to-normal vision. All of them had limited exposure to Italian, had visited Italy for no more than a month, and had at least 6 years of prior experience in learning English. The cohort was recruited *via* the Institute of Active Aging,

Hong Kong Polytechnic University. Hong Kong version of Montreal Cognitive Assessment (Wong et al., 2009) was used to confirm that they were cognitively normal (mean = 27.9, $SD = 1.7$). The attendance for the intensive Italian learning programme was high, with 21 of 25 participants completing all the 21 possible visits. There were four drop-outs midway through the programme, including three who completed 5 visits and one who completed 10 visits. One female participant was left-handed (handedness = -60) according to the Edinburgh inventory (Oldfield, 1971); only the remaining 20 participants were included in the present analysis. Their mean age, education, and HK-MoCA were 63.7 ($SD = 2.9$, range = 58.6–69.3), 15.2 ($SD = 3.1$, range = 11–21), and 27.8 ($SD = 1.7$, range = 25–30), respectively. All procedures were approved by the Ethical Review Committee, Hong Kong Polytechnic University. Written signed informed consent was obtained from all participants. They were reimbursed HKD 800 for their participation.

2.2. Study Description

The participants individually attended an intensive Italian vocabulary learning programme (**Figure 1**). The whole programme comprised an introductory lesson (visit 1), ten computerized vocabulary lessons (visits 3–4, 6–7, 9–10, 12–13, and 15–16), four revisions sessions (visit 8, 11, 14, 17), an Italian final test (visit 18), and a range of tests arranged at various points of the programme. These included an MRI session (visit 2), initial-learning phase cognitive tests (visit 5), initial-learning phase phonological tests (at the beginning of visits 3, 4, and 6), post-learning phase phonological tests (visits 18–19), post-learning phase cognitive tests (visit 20), and a debriefing session (visit 21; see **Supplementary Section S1**). The MRI, phonological, and cognitive testing were not conducted right at the beginning, for two reasons. (1) In reality, many language learners have some degree of knowledge of the target foreign language before proceeding to take a formal course. The introductory session serves to provide a more uniform experience prior to the MRI visit. Also, the introductory session gave them a taste of learning a new foreign language and offered a buffer period for them to commit to participating in the structural MRI session and the subsequent intensive Italian learning programme, which would likely reduce the dropout rate. (2) Upon enrolling to the study, most participants were anxious and most motivated to start learning Italian. As such, arranging too many tests that are unrelated to Italian learning right at the beginning would likely be seen as non-ideal arrangement from the learner's view, which might increase the dropout rate. Hence, the phonological and cognitive testing were not arranged immediately after the MRI session, but they were spread over sessions 3–5 as a compromise, especially because it was not our primary goals to investigate learning-induced cognitive gains. Overall, the whole programme can be divided into three phases—initial learning (visits 1–5), intensive-learning (visits 6–17), and post-learning (visits 18–21). The division of visits 1–17 into two phases was an arbitrary one, but the main difference was that a variety of tests were throughout the initial learning stage, while a revision session was arranged every two lessons only during the intensive-learning phases. Also, there was more separation between visits in the initial-learning phase

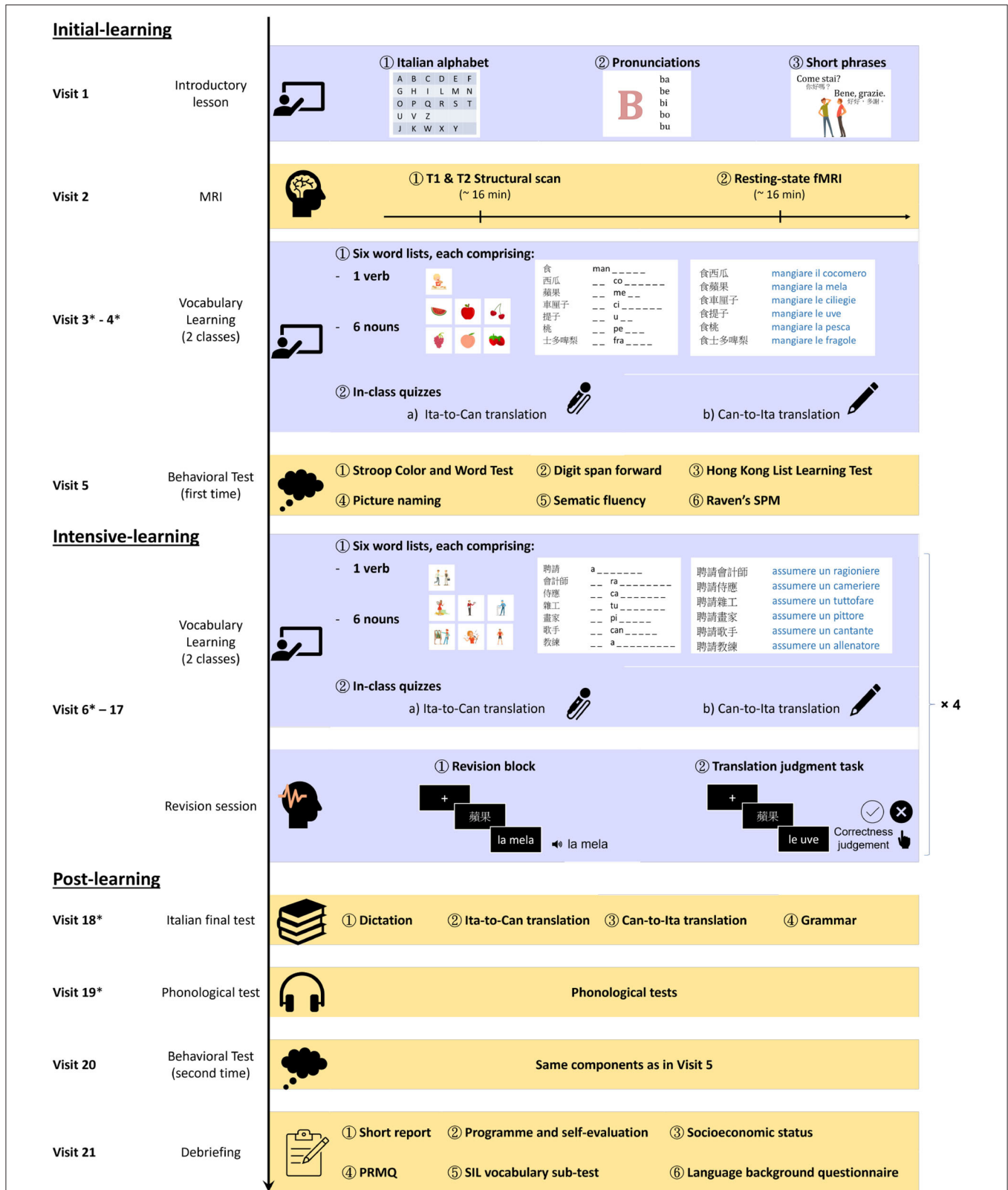


FIGURE 1 | Italian learning programme. The whole programme was divided into three phases—initial-learning (visits 1–5), intensive-learning (visits 6–17), and post-learning (visits 18–21). In the initial- and intensive-learning phases, six word lists were taught per vocabulary lesson, each comprising a verb and six nouns. After two vocabulary lessons, there was a revision session. Such a cycle of three sessions was repeated four times during the learning programme (hence $\times 4$ in the figure). Asterisks (*) were used to indicate the visits during which phonological tests were administered (see main text).

(4.5 days on average, $SD = 1.1$ days) than in intensive-learning (3.2 days, $SD = 0.7$ days), due to scheduling difficulties. The completion time of the first twenty visits of the programme was 68.9 days on average ($SD = 11.6$ days). The details for each component are laid out below in the following sections: Italian vocabulary learning programme, MRI protocol, and behavioral test battery.

2.3. Italian Vocabulary Learning Programme

2.3.1. Introductory Lesson

Each participant attended a 1.5-h introductory Italian lesson, in which they watched a custom-made video under the direction of an experimenter. There were three main parts in the video: Italian alphabet, Italian sounds, and Italian short phrases. Sixteen syllable identification questions were given as homework.

2.3.2. Computerized Vocabulary Lessons

Each computerized vocabulary learning lesson was about 120 min long. The goal of each lesson was to learn six word lists, each comprising a transitive verb (e.g., *mangiare*, “to eat”) and six matching nouns (e.g., *la mela*, “an apple.”) Thus, over the course of ten vocabulary lessons, there were 60 word lists in total, comprising 60 verbs and 360 nouns (see **Supplementary Table 1**, for the complete list of Italian words used, along with their Cantonese and English translation). In constructing these word lists, control measures were taken so that any systematic changes in the learning performance over time would not be due to confounding factors like the inherent difficulties of the Italian word lists or their Cantonese counterparts. In particular, because Italian words with a larger number of syllables or letters would likely be more difficult to remember, 5 groups of 12 word lists were constructed, with each group being the study materials for two consecutive classes. Across all pairs of groups, the verbs and nouns were matched separately in number of syllables (verbs: $ps > 0.14$, nouns: $ps > 0.10$) and letters (verbs: $ps > 0.32$; nouns: $ps > 0.14$). The Cantonese translation of the word list were also matched in the number of Chinese characters (verbs: $ps > 0.10$; nouns: $ps > 0.07$) and strokes (verbs: $ps > 0.28$; nouns: $ps > 0.12$). The stimulus matching was not conducted by lesson, because, due to the non-arbitrary pairing of the verbs and nouns, it was exceedingly difficult to create 10 groups that are matched in all four variables. Dividing by week also provided a better match in terms of the structure of the whole programme, in that, during the intensive-learning phase, the vast majority of older adults had two lessons per week, and a revision session was arranged every two lessons. By using week rather than lesson as a measurement unit, each in-class score is the average over two lessons, which should have lower measurement noise compared to that derived using only one lesson. For more details about the stimulus matching and construction, see **Supplementary Section S1** and **Supplementary Table S1**. The teaching materials were delivered based on a custom-made E-Prime 3.0 script.

For each word list, the participant first learnt the verb through a picture depicting the verb and by listening to its pronunciation three times. Next, the participant learnt the six nouns through clicking on the corresponding pictures to listen to their pronunciations. Each picture could be clicked three times,

before it would vanish on-screen. The order and pace of learning were self-determined. The participant was then given a chance to read aloud the seven words, upon hearing the pronunciation of each word. Next, they completed a fill-in-the-blank exercise on paper, in which they should spell all the seven words learnt. They were asked to check their own answer directly afterwards. The participant then listened to the pronunciation of the six verb-noun phrases (e.g., *mangiare la mela*, meaning “to eat an apple”) for three times each, before continuing with the next word list. The learning of each list usually took 15–20 min. After learning all six word lists, a short quiz was given at the end of each lesson. They were given as much time as they wish before the short quiz. The fill-in-the-blank and the quiz were included to promote memory consolidation and to assess immediate learning success, respectively.

There were two parts in the quiz: Italian-to-Cantonese and Cantonese-to-Italian translation (*abbr.* Ita-to-Can and Can-to-Ita thereafter). In either part, all 36 phrases learnt were tested, with each verb and noun would be tested six times or once, respectively. In Ita-to-Can, the participants read a verb-noun phrase (e.g., *mangiare la mela*) on screen and listened to its pronunciation, and they were asked to orally translate the phrase into Cantonese. The maximum score for this part was $36 \times 2 = 72$. In Can-to-Ita, the participants were given a paper-and-pen test, and they should write down the Italian translation of Cantonese verb-noun phrases. The order of the phrases was pseudorandomized, with no more than two consecutive phrases sharing the same verb. No feedback about the correct answer was given. For simplicity, participants were asked to write down the taught form of article (definite or indefinite) for the nouns in Can-to-Ita. However, the article was not scored because there was more than one possible grammatically correct choices, and their distinction was only learnt gradually; thus the maximum score remained as 72. As such, the grammatical use of the article was only scored in the final test.

In addition, six pages of handouts (one word list per page) showing the concepts learnt in pictorial form were given at the end of each lesson, on which the participants should copy the Italian words, either immediately or at home. Though the lexical consolidation process would be influenced by the number of times that participants revised the Italian vocabularies, the notes were essential for the effectiveness of the lessons. Also, although these lessons put more emphasis on vocabulary building, basic grammatical components and rules (including gender, singular and plural forms, article, present tense, adjective, numerals, and possessive adjective) were also taught, through four handouts and corresponding homework exercises, distributed at the end of lesson 1, 6, 8, and 10. Participants attended four revision sessions in total, each time following the completion of two vocabulary sessions. Results for the revision sessions, which involved electroencephalogram recordings, will be reported in a separate paper (under preparation).

2.3.3. Italian Final Test

An 1.5-h Italian test comprising four different parts—dictation, Ita-to-Can, Can-to-Ita, and grammar—was administered about one week after the final vocabulary lesson. In the first three parts, the words were uniformly sampled across the ten vocabulary

learning lessons. The first three parts were delivered using E-Prime 3.0, while the grammar test was an open-note, paper-and-pen test. For dictation, 60 nouns (including the article, e.g., *il gesso*, “the chalk”) were presented aurally at a rate of 15 s / item. Participants were asked to dictate each word on an answer sheet. The nouns would be repeated at a rate of 5 s / item, during which they could check and change their answers. Each correct spelling of the article and noun was worth half a point. The procedures for Ita-to-Can and Can-to-Ita translation were the same as those for the vocabulary lessons, except that there were only 30 verb-noun phrases here. While there were 30 questions in each test, the maximum score of Ita-to-Can was $30 \times 2 = 60$ because participants were asked to translate one Italian verb-noun phrase into one Cantonese verb and one Cantonese noun; the Cantonese verb and noun accounted for a point each. The maximum score of Can-to-Ita was $30 \times 3 = 90$ because participants were asked to translate one Cantonese verb-noun phrase into a verb, an article, plus a noun. For the grammar test, the questions were multiple-choice questions with either two or four choices; correspondingly, each question was worth two or four marks. The distribution of points approximately matched the relative emphasis in the grammatical notes. For the details of the final test, see **Supplementary Section S1** and **Supplementary Table S2**.

2.3.4. Measures of Immediate Learning Success and Language Retention

The participants' performance in each lesson was summarized by means of two scores: Ita-to-Can translation (maximum: 42) and Can-to-Ita translation (maximum: 42). For the final test, the maximum for dictation, Ita-to-Can, Can-to-Ita, and grammar were 60, 60, 90, and 180, respectively. Partial credit was not given; the Italian spellings or Cantonese words should be produced entirely.

2.4. MRI Protocol

2.4.1. Structural MRI

The MRI session was arranged after the introductory lesson. All patients were scanned with a Signa Premier 3T scanner (GE Healthcare, USA), located at the MRI Unit, Department of Radiology, University of Hong Kong. A head coil with forty-eight channels was used. High-resolution T1-weighted structural scans were acquired using a T1-weighted sequence known as 3D Gradient-Echo BRAin VOlume (BRAVO; TR 7.3 ms, TE 3 ms, TI 900 ms, flip angle 8° , $1 \times 1 \times 1 \text{ mm}^3$ voxels, TA = 5 min 42 s, FOV = $256 \times 256 \times 376 \text{ mm}^3$). T2 and resting-state fMRI data were also acquired, but they were not analyzed for the present paper. The total acquisition time was about 35 min.

2.4.2. Extraction of Anatomical Measures

The T1 data were analyzed using FreeSurfer 7.2.0 *recon-all* segmentation pipeline (Fischl, 2012). The automatic pipeline started with various pre-processing steps, including motion correction, skull-stripping, removal of non-brain tissues, and intensity normalization. The pipeline then proceeded with both surface-based morphometry (SBM) and volume-based morphometry.

Regarding the SBM part, a mesh model was constructed for the cortical surface, comprising an inner white matter surface that separated gray matter (GM) and white matter (WM) and an outer pial surface that separated GM and cerebrospinal fluid (Fischl and Dale, 2000). The surface models were then inflated, and co-registered to the *fsaverage* template. Desikan-Killiany cortical parcellation (*aparc.annot*) was adopted for cortical segmentation (Desikan et al., 2006). For each cortical ROI (**Figure 2**), four measures were extracted based on the surface models by summing/averaging across the vertices of the mesh: cortical thickness (the mean distance between the white and pial surface), cortical surface area (the sum of areas, measured at the white surface), cortical volume (the total volume of the gray matter that lies between the white and pial surface), and mean curvature (the average degree of cortical gyrification) (Winkler et al., 2010; Luders et al., 2012). Our rationale for using all four cortical measures was three-fold. (1) They have all been implicated in either language learning or cognition in general. (2) As a group, they would multivariately explain a larger percent of variance in the learning performance than they would individually, given that they likely encapsulate complementary information (e.g., Winkler et al., 2010; Yang et al., 2016). For example, there is only a weak correlation between cortical thickness and cortical surface area (Winkler et al., 2010). (3) These measures also have different developmental trajectories across the lifespan (Hogstrom et al., 2013), differential associations with cognitive functions (Gautam et al., 2015; Chung et al., 2017; Green et al., 2018; Tadayan et al., 2020) or even genetic factors (Grasby et al., 2020). For example, not only were individual differences in cortical thickness and cortical surface area ascribed to largely different genetic factors, but these factors also exert their influences in different developmental stages (Grasby et al., 2020).

The volume-based morphometry involves a non-linear volumetric registration to the FreeSurfer atlas, with the structural labeling being performed using a Gaussian Classifier Atlas (GCA; Fischl et al., 2002). For each sub-cortical ROI (**Figure 2**), the sub-cortical volume was extracted as the single measure.

2.5. Behavioral Test Battery

2.5.1. Cognitive Test Battery

A neuropsychological test battery comprising six standard cognitive tests was conducted, with the language of instruction being the native language of the participants (Cantonese). These tests included the Stroop Color and Word Test (Golden and Freshwater, 1978; Fong et al., 2020), digit span forward, Hong Kong List Learning Test (HKLLT; Chan and Kwok, 1998), picture naming (Bates et al., 2003; Fong et al., 2020), semantic fluency (Fong et al., 2021), and Raven's Standard Progressive Matrices (Raven and Court, 1998), and they were administered after three Italian learning lessons and one week post-learning. These tests assess processing speed, inhibition, short-term memory, verbal episodic memory, semantic knowledge, and matrix reasoning, respectively (**Table 1**). This battery was administered at the fifth visit (initial-learning phase) and the twentieth visit (post-learning phase) to assess the learning-induced cognitive gains, except for the possible use of two

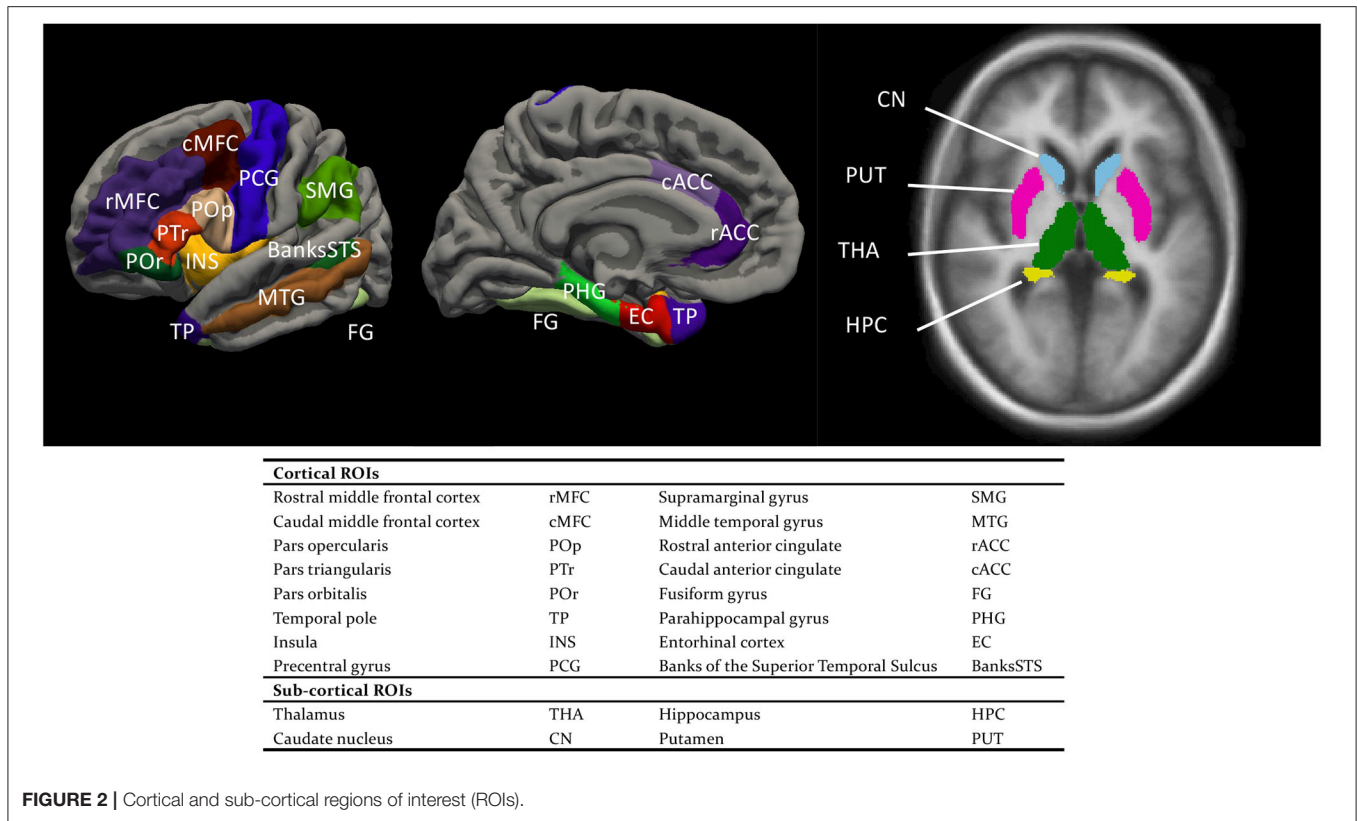


FIGURE 2 | Cortical and sub-cortical regions of interest (ROIs).

TABLE 1 | Measures of behavioral test performance.

Task	Cognitive function or phonological skill	Detail
Stroop Color and Word Test	Processing speed	Average number of correct responses in 45 s across the word and color sub-tasks
	Inhibition function	Number of correct responses in the color-word sub task, divided by the processing speed measure above
Digit span forward	Phonological STM	The maximum memory load reached (test continued if 2 of 3 questions were correct at a given load; half a point would be added in case of a single correct answer)
Hong Kong List Learning Test	Verbal episodic memory	Number of memory items retrieved after 30 min
Picture naming task	Naming latency	The average latency of correct answers
Semantic fluency	Semantic memory	Average number of responses across 16 semantic categories
Raven's SPM	Matrix reasoning	Standardized number of correct responses
Cantonese discrimination	Cantonese perceptual skill	Number of correct answers (Max: 41)
Cantonese Spoonerism	Cantonese phonological awareness	Number of correct answers (Max: 30)
English discrimination	English perceptual skill	Number of correct answers (Max: 38)
English Spoonerism	English phonological awareness	Number of correct answers (Max: 30)
Italian discrimination	Italian perceptual skill	Number of correct answers (Max: 35)
Italian Spoonerism	Italian phonological awareness	Number of correct answers (Max: 30)

different versions of tests to minimize training effects. Details of this battery can be found in **Supplementary Section S2** and **Supplementary Tables S3–S5**.

2.5.2. Phonological Test Battery

Two phonological tests (discrimination and Spoonerism) each were run for all three languages—Cantonese (L1), English (L2), and Italian (L3). For the discrimination task, participants heard four auditory stimuli in succession in each trial (*c.f.* Koda,

1998). Three of the stimuli shared the same segmental or suprasegmental feature (in the case of Cantonese tone), with the fourth not sharing the feature in question. The number of blocks and trials for each language examined was different, due to the different number of distinctive segmental and suprasegmental feature in each language. The Cantonese discrimination task was divided into three blocks, in which the feature in question was initial consonant (25 trials), vowel (10 trials), and tone (6 trials), respectively. The English discrimination task was

divided into two blocks, in which the feature in question was initial consonant (26 trials) and vowel (12 trials). The Italian discrimination task was divided into two blocks, in which the feature in question was also initial consonant (25 trials) and vowel (10 trials). For Spoonerism, participants listened to a pair of words in succession. Their task was to swap the initial consonant of the two words. There were 15 trials in all three versions. The test would be terminated if the participants failed to score any point in three consecutive trials. The Cantonese and English tests were arranged on visits 3 and 4 (one per visit, order counterbalanced), and together in visit 19 (order counterbalanced). The Italian tests were arranged on visits 5 and 18. More details of this battery can be found in **Supplementary Section S3** and **Supplementary Tables S6–S9**.

2.6. Linear Mixed-Effects Modeling

Linear mixed-effects (LME) models were constructed to test our main hypothesis that semantic and episodic memory functions play a major role in language learning in older learners. First, baseline demographic models (baseline models thereafter) were constructed for each of the three scores (in-class Ita-to-Can, in-class Can-to-Ita, and final test). In each baseline model, in addition to either Week (for the in-class Ita-to-Can or Can-to-Ita) or TestPart (for the final test), three demographic variables (age, years of education, and gender) were included as fixed-effects predictors, while participants were included as a random-effects variable.

Next, two groups of analyses were conducted, each involving comparisons between the full model(s) of interest against the corresponding baseline models. The first analysis comprised both a cortical and a complementary sub-cortical analysis. In the cortical analysis, the surface-based morphological measures were used as predictors to test the neural associations with the in-class performances and the final test scores. The regions of interest (ROIs), illustrated in **Figure 2**, comprised the Broca's area (pars opercularis and pars triangularis), the regions supporting semantic memory (pars orbitalis, temporal pole, and middle temporal gyrus) and episodic memory (caudal and rostral middle frontal cortex, entorhinal cortex, and parahippocampal gyrus), as well as regions important for phonological processing (SMG, precentral gyrus, banks of superior temporal sulcus), procedural memory (insula), language switching (caudal and rostral anterior cingulate), and orthographical processing (fusiform gyrus). For each ROI, a full model was constructed that include all the predictors in the baseline model, four anatomical measures of the ROI, and the interactions between each anatomical measure with either Week or TestPart. Likelihood ratio tests were used to evaluate whether the full model of each ROI represented an improvement over the baseline model. Multiple comparisons were corrected using the false discovery rate (FDR) procedure. If a significant interaction was found, *post hoc* trend analysis was conducted to reveal how the degree of association of each anatomical measure with the language learning performance was modulated by Week or TestPart. In the sub-cortical analysis, the sub-cortical volumes were used to predict the in-class and final test scores. The eight regions of interest (ROIs) included the hippocampus, thalamus, caudate nucleus, and putamen on

TABLE 2 | Summary statistics of the scores for the in-class quizzes (Ita-to-Can and Can-to-Ita) and the final test.

Score (%)	Mean	SD	Min	Max
InClass-Ita-to-Can	71.8	20.4	26.7	93.5
InClass-Can-to-Ita	41.1	22.5	6.2	73.0
FinalTest-Dictation	59.5	17.1	30.0	91.7
FinalTest-Ita-to-Can	80.2	21.4	36.7	100
FinalTest-Can-to-Ita	56.4	27.7	16.7	96.7
FinalTest-Grammar	92.9	5.9	78.9	100

both hemispheres. For each ROI, a full model was constructed using the sub-cortical volume of the ROI and its interaction with either Week or TestPart. The subsequent likelihood ratio tests were FDR-corrected.

In the second group of analyses—cognitive analysis—the cognitive predictors, derived from the neuropsychological data, were included as the additional predictors for modeling the three scores (Bonferroni correction was applied). For each model, backward elimination was conducted to yield the final model.

2.7. Cognitive and Phonological Skill Comparisons Between Initial- and Post-learning

For completeness, we tested the changes in cognitive functions and phonological skills incurred between the initial- and post-learning, despite the lack of a control group to assess whether the changes were due to repetition effects or induced by language learning. For each cognitive/phonological measure, a pairwise Wilcoxon signed rank test (two-tailed) was conducted with FDR-correction.

3. RESULTS

3.1. Italian Learning Performances

Table 2 shows the summary statistics for the mean performance for the two in-class quizzes (Ita-to-Can and Can-to-Ita) and the final test, which were used as indices for immediate learning success and language retention, respectively. **Figure 3** illustrates the overall in-class learning performance across the five weeks; the verb and noun components were separately plotted to reflect the structure of the in-class quizzes, although only the overall learning performance was selected for statistical analysis. To test whether the improvement was significant across the five weeks, a linear mixed-effects model was fitted for each in-class quiz, using the demographic variables for predicting the performance (**Table 3**). In both cases, the main effect of Week was significant, $ps < 0.001$. For Ita-to-Can, *post hoc* pairwise comparisons with Tukey's correction showed that the performance for week 1 was significantly worse than all other weeks, while that for week 5 was significantly better than all other weeks, with no differences among Week 2–4 being significant. For Can-to-Ita, the *post hoc* pairwise comparisons

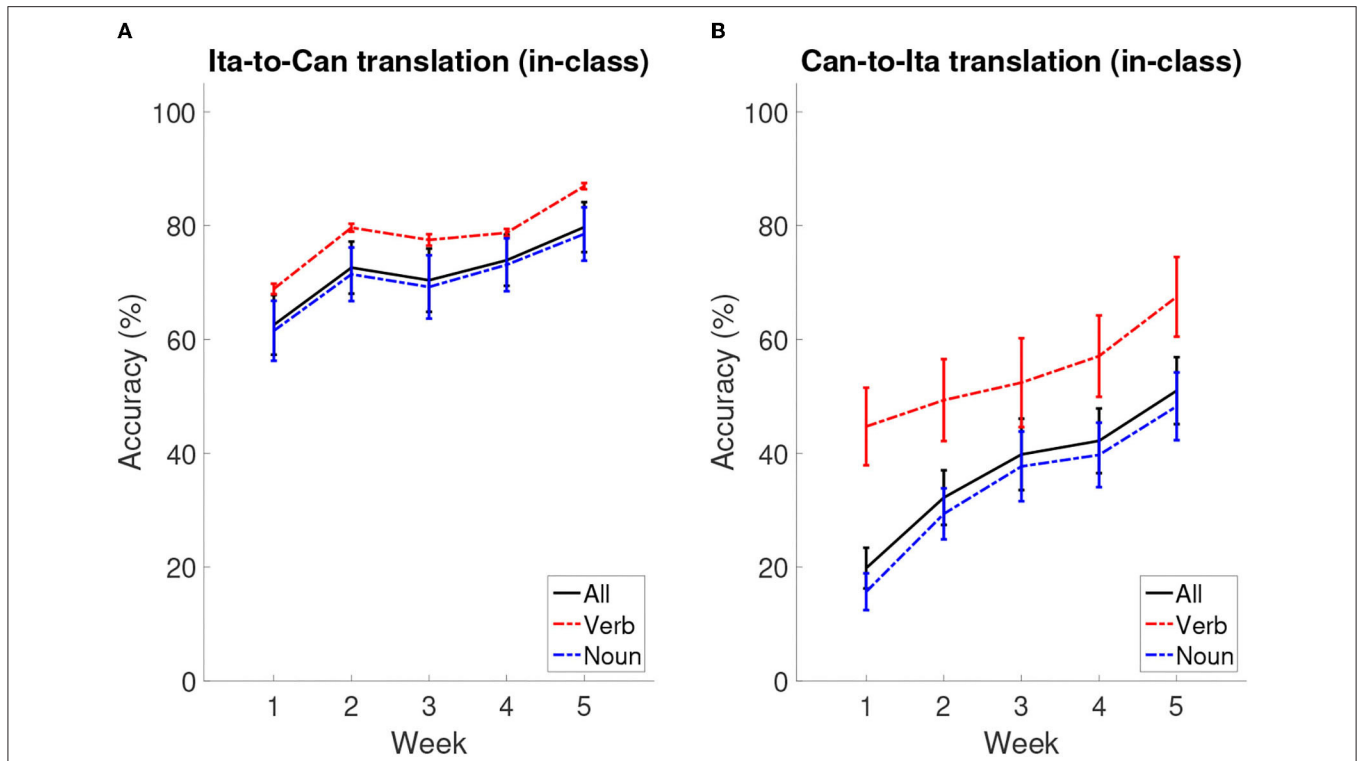


FIGURE 3 | Learning curves across the five weeks. **(A)** The accuracy of Ita-to-Can translation across weeks. **(B)** The accuracy of Can-to-Ita translation across weeks. Solid lines: overall in-class learning performance as a function of week; dotted lines: the scores for the verb and noun components, separately plotted to reflect the structure of the in-class quizzes. The error bars represent standard error means.

TABLE 3 | Demographic models for the in-class quizzes (Ita-to-Can and Can-to-Ita) and the final test.

Score	Term	SumSq	MeanSq	NumDF	DenDF	F	p	
Ita-to-Can	Week	3156.37	789.09	4	80.0	19.83	< 0.001	***
	Age	29.23	29.23	1	20.0	0.73	0.402	ns
	Education	25.39	25.39	1	20.0	0.64	0.434	ns
	Gender	21.36	21.36	1	20.0	0.54	0.472	ns
Can-to-Ita	Week	4726.65	1181.66	4	80.0	20.64	< 0.001	***
	Age	68.48	68.48	1	20.0	1.20	0.287	ns
	Education	43.20	43.20	1	20.0	0.75	0.395	ns
	Gender	0.59	0.59	1	20.0	0.01	0.920	ns
FinalTest	TestPart	18082.02	6027.34	3	60.0	41.41	< 0.001	***
	Age	307.16	307.16	1	20.0	2.11	0.162	ns
	Education	119.81	119.81	1	20.0	0.82	0.375	ns
	Gender	132.84	132.84	1	20.0	0.91	0.351	ns

***, $p < 0.001$, **, $p < 0.01$, *, $p < 0.05$, †, marginal, ns, non-significant.

indicated that the improvements over two weeks or more were all significant, although the improvement between successive weeks was only significant going from Week 1 to 2. These results suggested that the participants were generally able to learn an increasing number of Italian words across the five weeks.

For the final test, the participants were able to dictate 59.5% of the Italian words, backward translate 80.2% of Italian words

into Cantonese verbally, forward translate 56.4% of Cantonese words into Italian in written form, and achieve a score of 92.9% for grammar. It is worth noting that the first three parts were closed-note tests while participants were allowed to refer to the notes in the grammar part, which likely accounts for the higher grammar score. In the baseline model, the main effect of TestPart was significant, $p < 0.001$. *Post hoc* pairwise comparisons showed that the test score significantly

differed across the four parts, except between dictation and Can-to-Ita.

3.2. MRI Results

3.2.1. Surface-Based Cortical Measures and Brain Volume Measures

The four cortical anatomical measures (thickness, surface area, volume, and mean curvature) and the sub-cortical volumes were extracted for all cortical and sub-cortical ROIs (see **Supplementary Section S4** and **Supplementary Tables S10, S11**). In general, the cortical measures of the same ROI were strongly correlated, suggesting a high degree of redundancy in the measures. Despite the expected shared variance among the predictors, the associations with the learning performances could be best captured multivariately by the whole set of parameters. Therefore, all four predictors were included in testing the associations of the ROI with Italian learning performances. Special considerations were taken in interpreting the coefficients of the predictors (see section 3.2.2).

3.2.2. Cortical Predictors of Learning Performance

LME modeling was conducted on the in-class and final test scores. **Table 4** summarizes the cortical ROIs at which the anatomical parameters were sensitive to the learning performance (for the results at all the ROIs tested, see **Supplementary Tables S12, S14, and S16**). For transparency, the marginally significant associations (FDR-corrected $p < 0.10$) and all other associations with uncorrected $p < 0.05$ were also tabulated. Due to the strong correlations among the four predictors, sequential regression was applied to transform the original predictors into four adjusted predictors (more precisely, the most important predictor, determined by the chi-square test, was kept unchanged; in other words, only three predictors were transformed). In this way, the same amount of variance was explained by these adjusted predictors and the original predictors, with the advantage of improved interpretability of the predictors. Type III sums of squares procedure was conducted to estimate each fixed-effect term, including the main effects and the four interaction terms (**Supplementary Tables S13, S15, and S17**). *Posthoc* trend analysis was conducted on the significant interactions using the Satterthwaite method for estimating degrees of freedom. For in-Class Ita-to-Can score, after FDR correction for multiple comparisons, the baseline model was only marginally improved by the anatomical measures of four ROIs: (1) left pars orbitalis, $p = 0.052$; (2) left caudal middle frontal cortex, $p = 0.075$; (3) right insula, $p = 0.075$; and (4) left entorhinal, $p = 0.082$ (see **Supplementary Section S5**, for the follow-up analyses on each ROI).

The prediction of in-class Can-to-Ita performance was significantly improved by the cortical anatomical measures of seven ROIs. For (1) left pars orbitalis, the performance was significantly associated with adjusted curvature, $t = 3.03$, $p = 0.007$, and adjusted area, $t = -2.35$, $p = 0.029$. Week \times adjusted area was significant, with the negative association being significant only for Week 3–5, and that the association was significantly larger in magnitude in Week 5 than both Weeks 1 and 2 (**Figure 4A**). Week \times adjusted volume was also significant.

Although the association was non-significant in all weeks, there were subtle pairwise differences in the strength of association across different weeks. For (2) right pars orbitalis, Week \times adjusted volume was significant, but the *post hoc* trend analysis revealed no significant association. Like left pars orbitalis, the interaction was due to subtle pairwise differences in the strength of association across different weeks. For conciseness, only significant associations were reported thereafter. For (3) right insula, the performance was significantly predicted by adjusted volume, $t = 4.10$, $p < 0.001$, adjusted area, $t = -3.95$, $p < 0.001$, and adjusted curvature, $t = -2.58$, $p = 0.018$. Week \times adjusted volume was significant, with the association being significant in all five weeks; the interaction was due to the significantly larger association strength in the final two weeks (Weeks 4 and 5) than Week 2. For (4) right caudal anterior cingulate cortex, both Week \times adjusted area and Week \times adjusted curvature were significant, but none of the association was significant in any of the weeks. For (5) left entorhinal cortex, the performance was significantly associated with adjusted thickness, $t = 3.62$, $p = 0.001$, and adjusted curvature, $t = -5.16$, $p < 0.001$. There was a significant Week \times adjusted curvature, which was due to the much more negative association in Week 5 than in both Week 1 and 2. For (6) left caudal middle frontal cortex, there was significant association between Can-to-Ita performance with adjusted curvature $t = 2.75$, $p = 0.013$, and with adjusted thickness $t = 2.48$, $p = 0.022$. Week \times adjusted thickness was significant, which was attributed to the fact that the association was significant only for Week 3–5. For (7) right middle temporal gyrus, Week significantly interacted with adjusted area and adjusted volume. However, the association between performance and neither measure was significant in any of the weeks. For an understanding about the subtle changes in association across weeks, readers may refer to **Figure 4B**. Apart from the seven ROIs above, the baseline model was marginally improved at five additional ROIs (left pars triangularis, left temporal pole, right pars triangularis, right entorhinal cortex, and right banks of STS).

For final test, significant associations of the test score with the cortical measures were found for nine ROIs. For (1) right entorhinal cortex, there was significant positive association between final test score and adjusted curvature, $t = 3.06$, $p = 0.006$, adjusted thickness, $t = 2.56$, $p = 0.019$, and adjusted area, $t = 2.12$, $p = 0.047$. TestPart \times adjusted volume was significant, as the adjusted volume was negatively associated with only the two translation scores (Ita-to-Can, and Can-to-Ita). TestPart also interacted with adjusted curvature, which was positively associated with the first three parts (Dictation, Ita-to-Can, and Can-to-Ita), and adjusted thickness, which had the same association pattern except that its association with Dictation was only marginally significant. For (2) left pars orbitalis, the test score was significantly predicted by adjusted volume, $t = 3.15$, $p = 0.005$. The interaction between TestPart and adjusted volume was also significant, meaning that the association between the test score and adjusted area was modulated across different parts. For (3) left banks of STS, the test score was significantly and negatively associated with adjusted thickness, $t = -2.63$, $p = 0.016$. There were significant interactions

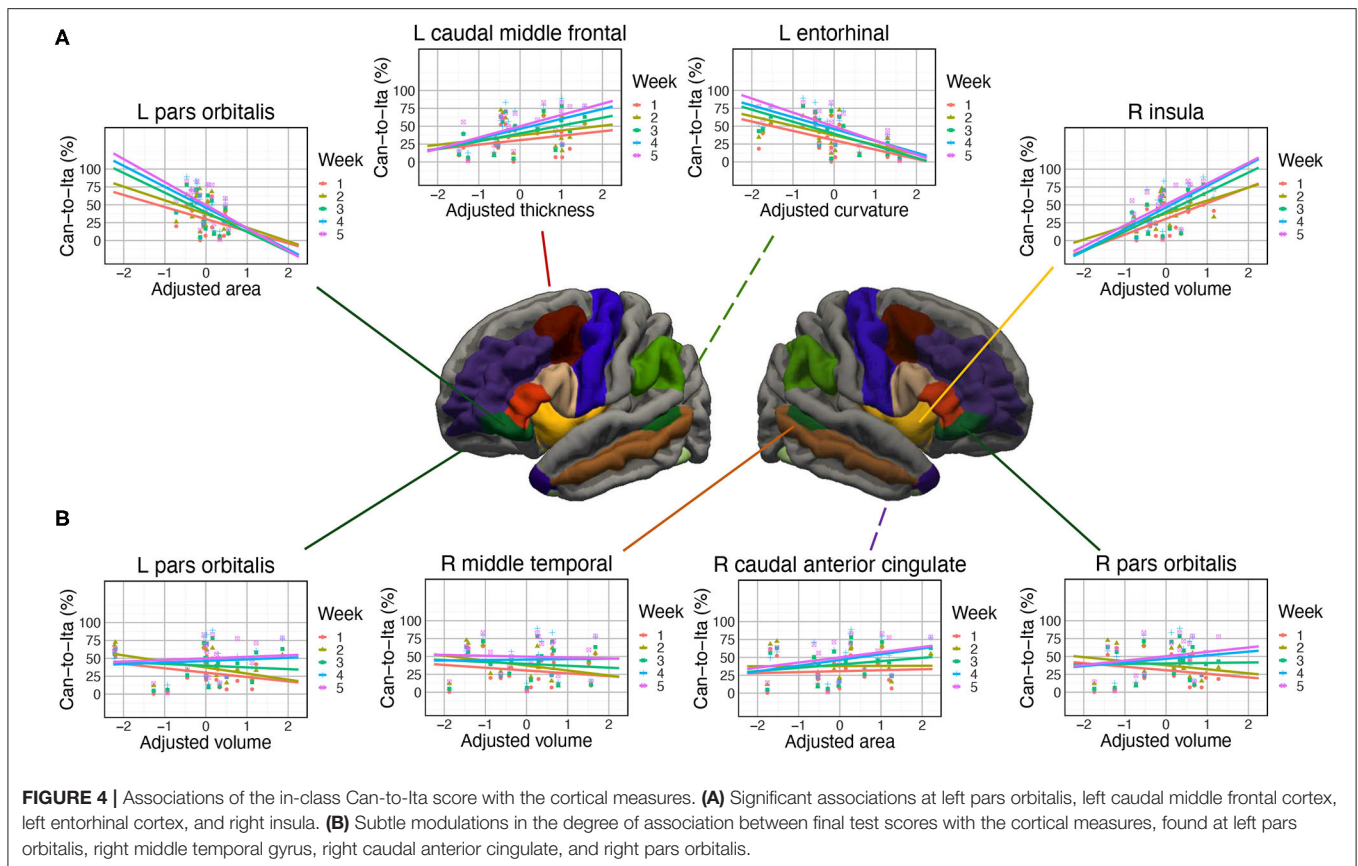
TABLE 4 | Cortical ROIs showing significant associations with the Italian learning performance.

Score	Region	AIC	BIC	logLik	r2m	r2c	χ^2	df	p (unc)	p (FDR)	
Ita-to-Can	L pars orbitalis	742.99	821.14	-341.49	0.67	0.94	43.75	20	0.002	0.052	†
	R insula	747.47	825.63	-343.74	0.75	0.93	39.26	20	0.006	0.075	†
	L caudalmiddlefrontal	747.91	826.07	-343.96	0.63	0.94	38.82	20	0.007	0.075	†
	L entorhinal	749.27	827.42	-344.63	0.58	0.94	37.46	20	0.010	0.082	†
	R parsorbitalis	751.16	829.31	-345.58	0.33	0.94	35.58	20	0.017	0.110	ns
	L parstriangularis	751.90	830.06	-345.95	0.59	0.93	34.83	20	0.021	0.112	ns
	L middletemporal	754.23	832.38	-347.11	0.49	0.94	32.51	20	0.038	0.163	ns
	R rostralmiddlefrontal	754.50	832.65	-347.25	0.53	0.93	32.23	20	0.041	0.163	ns
	R entorhinal	755.31	833.46	-347.65	0.42	0.94	31.43	20	0.050	0.177	ns
Can-to-Ita	L parsorbitalis	739.46	817.62	-339.73	0.60	0.96	79.54	20	< 0.001	< 0.001	***
	R parsorbitalis	773.79	851.94	-356.89	0.36	0.94	45.22	20	0.001	0.016	*
	R insula	775.95	854.10	-357.97	0.72	0.92	43.05	20	0.002	0.019	*
	R caudalanteriorcingulate	776.55	854.71	-358.28	0.32	0.94	42.45	20	0.002	0.019	*
	L entorhinal	778.46	856.61	-359.23	0.70	0.92	40.55	20	0.004	0.027	*
	L caudalmiddlefrontal	779.93	858.08	-359.96	0.56	0.93	39.08	20	0.007	0.035	*
	R middletemporal	781.39	859.55	-360.70	0.33	0.94	37.61	20	0.010	0.045	*
	L parstriangularis	782.35	860.51	-361.18	0.47	0.93	36.65	20	0.013	0.052	†
	L temporalpole	782.89	861.04	-361.44	0.64	0.92	36.11	20	0.015	0.053	†
	R parstriangularis	784.01	862.16	-362.00	0.30	0.94	34.99	20	0.020	0.064	†
	R entorhinal	785.88	864.03	-362.94	0.49	0.93	33.13	20	0.033	0.089	†
	R bankssts	785.97	864.12	-362.98	0.41	0.93	33.04	20	0.033	0.089	†
Final test	R entorhinal	659.81	719.36	-304.91	0.75	0.86	48.06	16	< 0.001	0.001	**
	L parsorbitalis	665.77	725.32	-307.89	0.71	0.86	42.10	16	< 0.001	0.006	**
	L bankssts	668.01	727.57	-309.01	0.69	0.86	39.86	16	0.001	0.009	**
	R fusiform	668.87	728.42	-309.43	0.74	0.83	39.01	16	0.001	0.009	**
	L supramarginal	672.66	732.22	-311.33	0.70	0.84	35.21	16	0.004	0.024	*
	L precentral	673.41	732.96	-311.70	0.67	0.85	34.47	16	0.005	0.025	*
	L parstriangularis	674.62	734.17	-312.31	0.67	0.84	33.26	16	0.007	0.031	*
	L rostralmiddlefrontal	676.83	736.38	-313.41	0.68	0.83	31.05	16	0.013	0.048	*
	L caudalmiddlefrontal	676.87	736.43	-313.44	0.70	0.82	31.00	16	0.013	0.048	*
	R supramarginal	677.77	737.32	-313.89	0.65	0.84	30.10	16	0.017	0.056	†

p(unc), uncorrected p, p(FDR), FDR-corrected p. ***, $p < 0.001$, **, $p < 0.01$, *, $p < 0.05$, †, marginal, ns, non-significant.

of TestPart with adjusted thickness (which was significantly and negatively associated with translation scores) and adjusted area (which was marginally and negatively associated with both translation scores). For (4) right fusiform gyrus, the test score was significantly predicted by adjusted thickness, $t = -4.59$, $p < 0.001$. TestPart \times adjusted thickness was significant: adjusted thickness was negatively associated with the scores in the first three parts. For (5) left SMG, the test score was positively associated with adjusted volume, $t = 2.21$, $p = 0.039$, and negatively associated with adjusted area, $t = -2.33$, $p = 0.030$. TestPart \times adjusted area was also significant, with the negative association being significant for both translation scores. For (6) left precentral gyrus, TestPart interacted with both adjusted thickness and adjusted curvature, with the thickness being significantly associated with the two translation scores, while adjusted curvature was only significantly associated with Can-to-Ita. For (7) left pars triangularis, the test score was significantly predicted by adjusted volume, $t = 2.52$, $p =$

0.021. TestPart \times adjusted volume and TestPart \times adjusted curvature were significant, with both adjusted volume and curvature being positively contributing to the two translation scores. For (8) left caudal middle frontal cortex, the test score was significantly predicted by adjusted curvature, $t = 3.16$, $p = 0.005$. TestPart also interacted with adjusted curvature and adjusted area, with the adjusted curvature being positively associated with the test scores for the first three parts, and the adjusted area negatively associated with the two translation scores. For (9) left rostral middle frontal cortex, the test score was significantly and positively associated with adjusted volume, $t = 2.19$, $p = 0.041$, and with adjusted curvature, $t = 2.11$, $p = 0.047$. TestPart \times adjusted curvature was significant, with the curvature being positively associated with the two translation scores. Marginal improvement over the baseline model was observed at one additional ROI (right SMG). For each ROI above, the representative interaction term was illustrated in **Figure 5**.



3.2.3. Sub-cortical Volumetric Predictors of Learning Performance

The contributions of sub-cortical regions were also examined using sub-cortical volumes, normalized by the intracranial volume (ICV). None of the normalized sub-cortical volumes could significantly improve the two baseline models for the in-class vocabulary learning performances: Ita-to-Can, all uncorrected p 's > 0.13, Can-to-Ita, all uncorrected p 's > 0.16. For the final test score, the baseline model was significantly improved by two predictors, namely, the sub-cortical volume of left and right thalamus (Table 5). For left thalamus, the interaction TestPart \times normalized volume was significant, $F(3, 60) = 4.29$, $p = 0.008$, due to the positive association of normalized volume with the scores in two parts: Ita-to-Can, $t = 2.24$, $p = 0.031$, and Can-to-Ita, $t = 2.35$, $p = 0.024$. For right thalamus, the interaction TestPart \times normalized volume was significant, $F(3, 60) = 3.74$, $p = 0.016$, due to the significant positive associations of normalized volume with Can-to-Ita, $t = 2.73$, $p = 0.009$ and Ita-to-Can, $t = 2.43$, $p = 0.019$. There was only a weak indication for an association with the sub-cortical volume of right hippocampus (uncorrected $p = 0.09$).

3.3. Cognitive and Phonological Associations

3.3.1. Correlations Among the Cognitive and Phonological Predictors

Figure 6 shows the correlations among the cognitive and phonological predictors. In general, there were high correlations

among the three phonological discrimination scores (average $r = 0.58$, range = 0.45 – 0.76) and among the three Spoonerism scores (average $r = 0.52$, range = 0.47 – 0.56), but relatively low correlations across these measures. This pattern of correlation supported our a priori choice of selecting only one measure each of phonological discrimination and of Spoonerism for representing phonological skills. Given the non-phonological nature of the Chinese writing system, the majority of the older learners were unfamiliar with the phonetic system used in Cantonese. Consequently, the English phonological discrimination and Spoonerism scores were selected as the phonological predictors.

3.3.2. Cognitive Models: Predictors of In-class and Final Test Performances

One of the three baseline models was significantly improved by the cognitive predictors: Can-to-Ita, $AIC = 789.83$, $BIC = 933.11$, $\logLik = -339.91$, $\chi^2 = 79.18$, $DF = 45$, uncorrected $p = 0.001$, corrected $p = 0.004$. Backward elimination was applied to yield the final model (Table 6). The performance was positively associated with HKLLT ($t(20) = 4.8$, $p < 0.001$), Raven's SPM ($t(20) = 4.33$, $p < 0.001$), English Spoonerism ($t(20) = 3.28$, $p = 0.004$), semantic fluency ($t(20) = 3.05$, $p = 0.006$), and processing speed ($t(20) = 2.49$, $p = 0.022$), and as expected, negatively associated with picture naming latency ($t(20) = -6.53$, $p < 0.001$), but it was also negatively associated with digit span forward ($t(20) = -6.34$, $p < 0.001$) and inhibition ($t(20) = -5.90$, $p < 0.001$). The positive

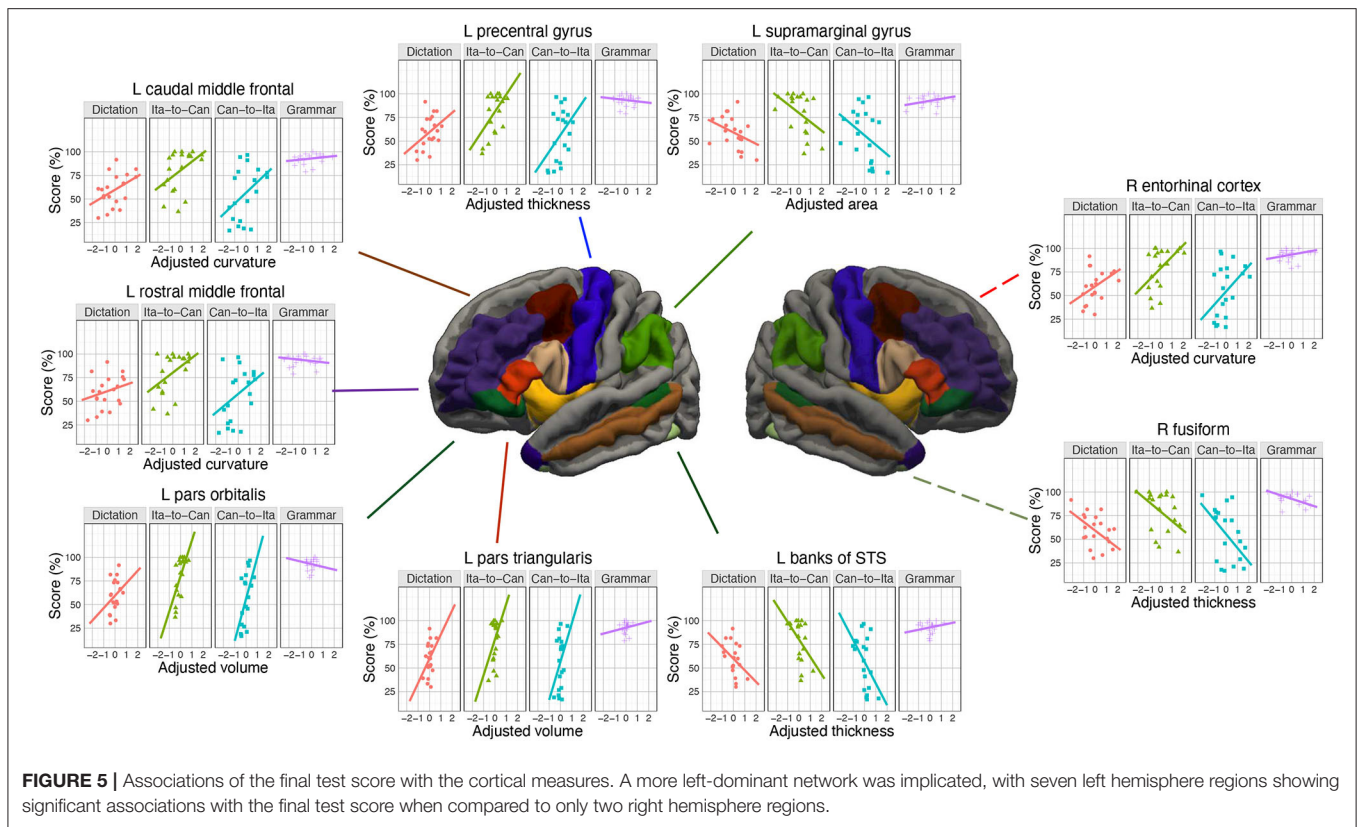


TABLE 5 | Sub-cortical ROIs showing significant associations with the final test scores.

Region	AIC	BIC	logLik	r2m	r2c	χ^2	Df	p (unc)	p (FDR)	
L thalamus	669.78	700.75	-321.89	0.58	0.80	14.09	4	0.007	0.036	*
L caudate	683.15	714.11	-328.57	0.51	0.76	0.73	4	0.948	0.948	ns
L putamen	680.32	711.29	-327.16	0.54	0.77	3.55	4	0.470	0.626	ns
L hippocampus	679.81	710.78	-326.91	0.53	0.77	4.06	4	0.397	0.626	ns
R thalamus	670.33	701.30	-322.17	0.59	0.80	13.54	4	0.009	0.036	*
R caudate	682.64	713.60	-328.32	0.53	0.76	1.24	4	0.872	0.948	ns
R putamen	679.95	710.91	-326.97	0.54	0.77	3.93	4	0.416	0.626	ns
R hippocampus	675.83	706.80	-324.92	0.56	0.78	8.04	4	0.090	0.240	ns

p(unc), uncorrected *p*, *p(fdr)*, FDR-corrected *p*.

***, *p* < 0.001, **, *p* < 0.01, *, *p* < 0.05, †, marginal, ns, non-significant.

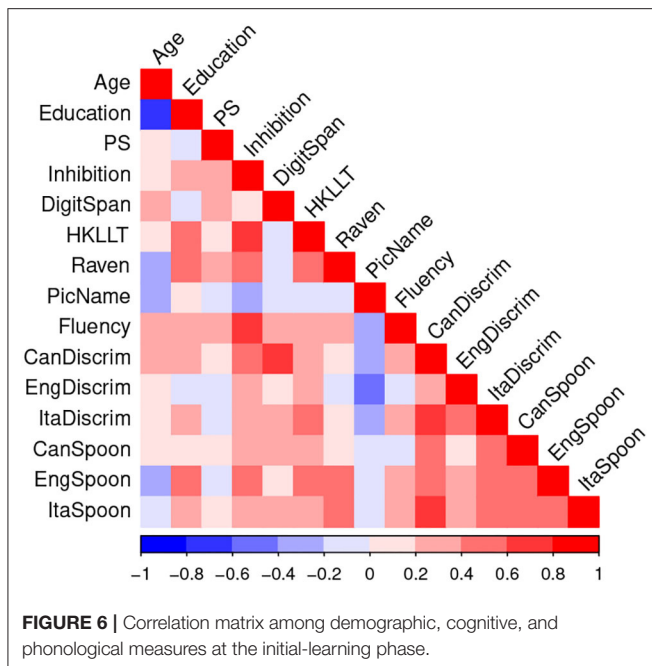
associations and the negative association with picture naming latency are illustrated in **Figures 7A–E**. In addition, Week × processing speed was also significant, $F(4, 80) = 4.07, p = 0.005$. This interaction effect was due to the general increase in the association of processing speed with the learning performance across time (**Figure 7F**), reaching statistical significance in Week 4 ($t(20) = 2.26, p = 0.029$) and Week 5 ($t(20) = 4.09, p < 0.001$).

For in-class Ita-to-Can, the improvement over the baseline model did not survive the correction, $AIC = 774.01, BIC = 917.30, logLik = -332.01, \chi^2 = 62.72, DF = 45$, uncorrected $p = 0.041$, corrected $p = 0.124$. The same held for the baseline model of the final test, $AIC = 696.15, BIC = 803.34$,

$logLik = -303.08, \chi^2 = 51.72, DF = 36, p = 0.043$. However, backward elimination was applied to explore the most important predictor(s). For the in-class Ita-to-Can, the exploratory analysis revealed a significant Lesson × HKLLT interaction; the trend analysis revealed marginally significant positive associations in week 1 ($t(80) = 2.05, p = 0.052$) and week 3 ($t(80) = 1.86, p = 0.076$). For the final test, it was positively associated with semantic fluency ($t(20) = 2.18, p = 0.041$).

3.4. Cognitive Comparison Between the Initial-Learning vs. Post-learning Phase

Pairwise Wilcoxon signed rank tests revealed marginally shorter picture naming latency, $p = 0.057$, with an effect size (King et al.,



2011) of $r_c = -0.67$, marginally better Raven score ($p = 0.057$, $r_c = 0.67$) and HKLLT ($p = 0.094$, $r_c = 0.67$). There were also marginally higher Italian discrimination score, $p = 0.057$, $r_c = 0.67$, as well as higher scores in the two Cantonese tests: CanDiscrim, $p = 0.094$, $r_c = 0.57$, and CanSpoon, $p = 0.094$, $r_c = 0.62$ (Table 7). However, due to the lack of a control group, it is unclear whether the change arose due to repetition effect or it was truly induced by intensive vocabulary learning.

4. DISCUSSION

4.1. Semantic, Episodic, and Phonological Associations With Vocabulary Learning

Taking an individual-differences approach, the present study used both anatomical and cognitive measurements to investigate the basis of FLL in older adults, focusing on whether semantic and episodic memory play an especially important role in vocabulary learning within this special group of language learners.

In support of the hypothesis that semantic functions are important for older language learners, the left pars orbitalis, which is central to semantic control (Sabb et al., 2007; Binder et al., 2009; Ralph et al., 2017) by virtue of its white matter pathway (uncinate fasciculus) to the semantic hub in the temporal pole (Harvey et al., 2013), was found to be associated with the performance in the in-class quiz and the final test. In particular, its adjusted curvature was positively associated with the in-class Can-to-Ita performance while its adjusted volume was significantly associated with the first three parts of the final test (dictation, Ita-to-Can, and Can-to-Ita). This suggests that high-performing learners tend to have higher adjusted curvature and volume in left pars orbitalis. To our knowledge,

the importance of the left pars orbitalis has seldom been reported in previous works in which the learner group comprised mostly younger adults, even though a whole-brain analysis was conducted in these studies (e.g., Bellander et al., 2016). Instead, it is generally understood that the anatomical measures of the left IFG, as the seat of the Broca's area, are sensitive to language processes and bilingualism. For example, its gray matter volume was positively correlated with L2 performance (Mårtensson et al., 2012; Stein et al., 2012). Its cortical thickness was also larger for bilinguals that acquired their L2 in late childhood (8–13 years) than in early childhood (4–7 years) (Klein et al., 2014). Our finding is consistent with this general understanding, but further suggests that the pars orbitalis is especially associated with both immediate and long-term vocabulary retention. It is worth noting that the left temporal pole has barely failed to reach statistical significance in predicting in-class Can-to-Ita (FDR-corrected $p = 0.053$), which would have provided another piece of convergent evidence.

In support of the hypothesis that episodic memory is also important for older language learners, associations with the performance of both the in-class quiz and the final test were found at the left caudal middle frontal cortex. In particular, the mean curvature of the left caudal frontal cortex was consistently positively associated with the learning performance, including the in-class Can-to-Ita score regardless of week and the first three parts of the final test. By comparison, the adjusted thickness only significantly predicted the in-class Can-to-Ita over Weeks 1–3 while the adjusted cortical surface area was negatively associated with the two translation scores in the final test. In the literature, the left caudal middle frontal cortex is known to be associated with semantic strategies (Kirchhoff et al., 2014; Yu et al., 2018), referential encoding strategies (Yang et al., 2013), self-initiated elaborative encoding strategies (Husa et al., 2017), and working memory (Petrides et al., 1993). The present finding suggests that better-performing older learners might have employed a variety of encoding strategies in vocabulary learning. The present result corroborated with two previous works. Yang et al. (2015) reported that the middle frontal cortex was more active in successful learners in a tone discrimination task. The authors suggested that the higher level of activation was associated with the path between middle frontal gyrus and inferior parietal lobe, where lexical knowledge was automatically activated and sent to middle frontal gyrus. Another study found that the cortical thickness of MFG was increased after FLL (Legault et al., 2019b). Apart from the left caudal middle frontal cortex, converging evidence was also found in the left entorhinal cortex, which is implicated in the episodic processing of object/event information in relation to the context (Eichenbaum et al., 2012). Specifically, the in-class Can-to-Ita performance was positively associated with its cortical thickness.

In contrast, for regions that are especially important for phonological processing, they were associated with only long-term retention. Specifically, the left precentral gyrus is important for phonological working memory, serving as a locus for phonological rehearsal (Veroude et al., 2010; Novén et al., 2019), while the left SMG is known to play important role in phonological processing skills (Veroude et al., 2010;

TABLE 6 | Cognitive models for the in-class Ita-to-Can and Can-to-Ita scores and for the final test, obtained after backward elimination of the full models.

Score	Term	SumSq	MeanSq	NumDF	DenDF	F	p	
Ita-to-Can ‡	Age	71.22	71.22	1	20.0	2.15	0.158	ns
	Gender	0.24	0.24	1	20.0	0.01	0.933	ns
	Education	2.23	2.23	1	20.0	0.07	0.798	ns
	Week	3156.37	789.09	4	80.0	23.87	< 0.001	***
	HKLLT	75.30	75.30	1	20.0	2.28	0.147	ns
	Week:HKLLT	539.17	134.79	4	80.0	4.08	0.005	**
Can-to-Ita	Age	139.41	139.41	1	20.0	2.93	0.102	ns
	Gender	410.31	410.31	1	20.0	8.62	0.008	**
	Education	19.29	19.29	1	20.0	0.41	0.532	ns
	Week	4726.65	1181.66	4	80.0	24.84	< 0.001	***
	PS	295.42	295.42	1	20.0	6.21	0.022	*
	Inhibition	1655.79	1655.79	1	20.0	34.80	< 0.001	***
	DigitSpan	1912.07	1912.07	1	20.0	40.19	< 0.001	***
	HKLLT	1095.99	1095.99	1	20.0	23.04	< 0.001	***
	PicName	2028.22	2028.22	1	20.0	42.63	< 0.001	***
	Fluency	441.17	441.17	1	20.0	9.27	0.006	**
	Raven	893.03	893.03	1	20.0	18.77	< 0.001	***
	EngSpoon	511.33	511.33	1	20.0	10.75	0.004	**
	Week:PS	773.63	193.41	4	80.0	4.07	0.005	**
	FinalTest ‡	Age	892.65	892.65	1	20.0	6.13	0.022
Gender		478.92	478.92	1	20.0	3.29	0.085	†
Education		6.57	6.57	1	20.0	0.05	0.834	ns
TestPart		18082.02	6027.34	3	60.0	41.41	< 0.001	***
Fluency		693.43	693.43	1	20.0	4.76	0.041	*

‡The likelihood ratio test comparing the baseline and full model of the in-class Ita-to-Can score was non-significant after Bonferroni correction; the result tabulated here was obtained to explore the cognitive predictors of the in-class Ita-to-Can performance (see main text).

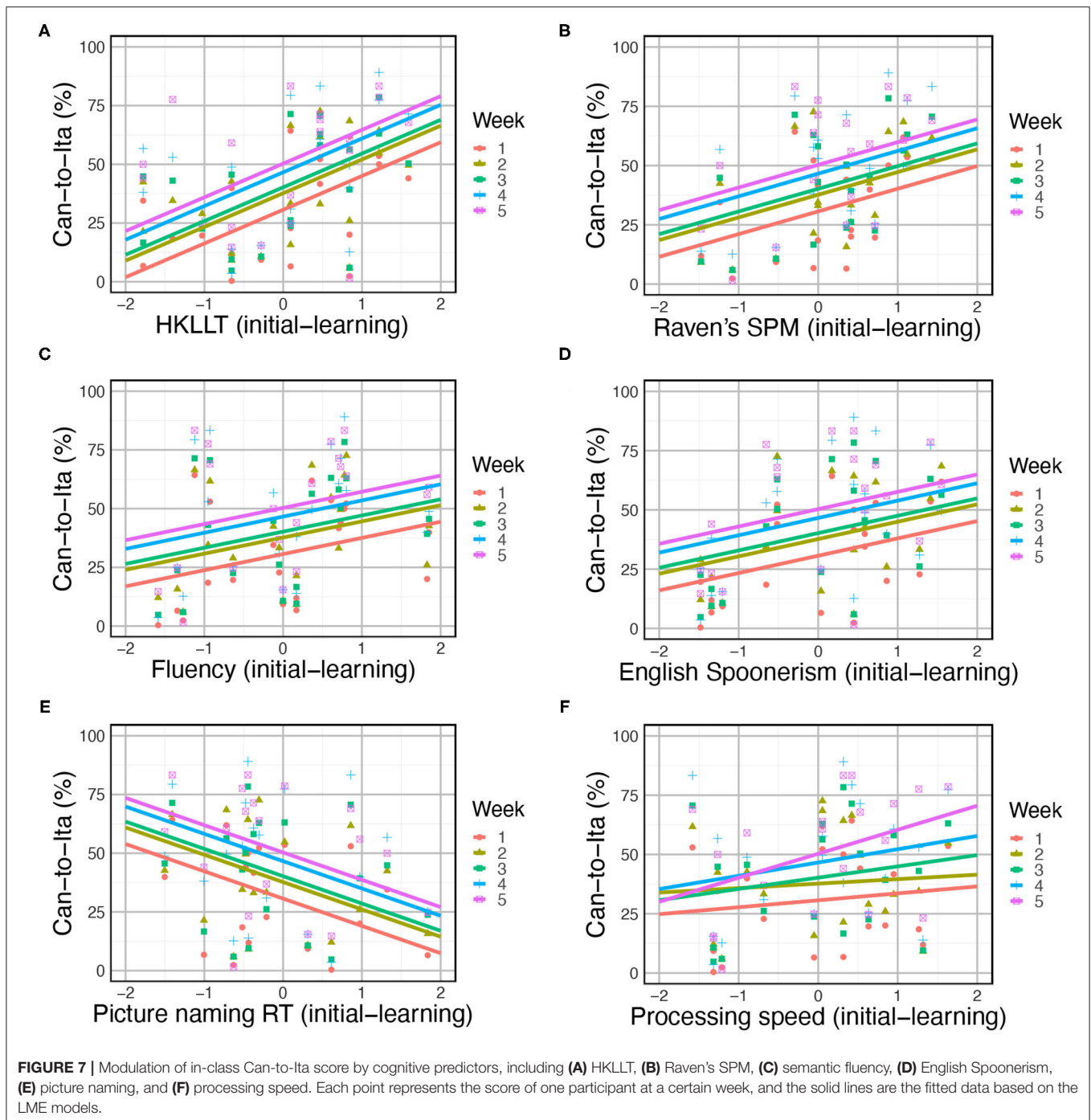
***, $p < 0.001$, **, $p < 0.01$, *, $p < 0.05$, †, marginal, ns, non-significant.

Sliwinska et al., 2012). In partial support to the hypothesis that phonological skills play a role in learning success, for left precentral gyrus, its adjusted thickness and curvature were significantly or marginally associated with the two translation tasks in the final test. The adjusted volume of left SMG was positively associated with the final test score in general, while the adjusted area was negatively associated with the scores in the two translation scores of the final test. The lack of associations of in-class scores with the precentral gyrus and left SMG associations, coupled with the presence of such associations with left pars orbitalis and left caudal middle frontal gyrus, is consistent with our hypothesis that the semantic and episodic memory functions play a more important role for FLL in older adults.

We speculate that semantic and episodic memory processes would be especially needed for intensive vocabulary learning; they are involved, for example, in strategically and proactively generating idiosyncratic semantics-based mnemonics (Thomas and Wang, 1996; Khoii and Sharififar, 2013) or referential encoding strategies for the new words (Turk et al., 2015). Suppose the learner is not confident that they can directly retrieve the meaning of *la mela* (“the apple,”) a possible cue for retrieval could be the semantic category “fruit,” which could in turn be encoded in terms of the learner’s native language. Referential encoding

strategies may also be relevant, as the learner could encode the contextual information, e.g., the noun “la mela” was learnt within the word list associated with the verb *mangiare* (“to eat”) or it was the first word in the first list. Having such cues facilitates the word recall during the in-class translation quizzes. Learners who used these cues may also tend to have better performance in long-term retention, even though the learners no longer need such strategies after many weeks of consolidation.

The cognitive associations reported are consistent with the putative reasons for the strong association of the FLL performance with the left pars orbitalis and left caudal middle frontal cortex discussed above. Across the three cognitive models, only semantic fluency and the delayed recall in the Hong Kong List Learning Test (HKLLT), as measures of semantic retrieval and verbal episodic memory functions, were significant predictors in more than one final models. Specifically, HKLLT significantly contributed to the two in-class scores but not the final test; this pattern suggested that verbal episodic memory is more associated with the immediate learning than long-term language retention performance. In contrast, semantic fluency significantly predicted both in-class Can-to-Ita and the final test. Thus, the evidence suggested that semantic fluency was associated with both immediate learning



and long-term retention. One reason that semantic fluency was not significant for predicting in-class Ita-to-Can is that controlled semantic retrieval ability, as captured by the semantic fluency task, has a strong association with the use of semantic associations and strategies. During vocabulary learning, most learners tend to be able to make a direct association of the L2 (Italian) word with the L1 (Cantonese) word, as reflected by a relatively high performance of the Ita-to-Can task in the present study. In contrast, they were much less accurate

in the Can-to-Ita task, indicating that they had difficulty in recalling the same association in the reverse direction, i.e., from the L1 (Cantonese) word to the L2 (Italian) word. Such asymmetry has been a point of emphasis in modeling the acquisition of foreign vocabulary, e.g., the parasitic model of vocabulary acquisition (Ecke, 2015). Due to the relative difficulty in Can-to-Ita, the task is subject to a greater use of mnemonic strategies for retrieval. The stronger association of semantic fluency with Can-to-Ita than with Ita-to-Can suggests that

TABLE 7 | Comparisons of cognitive and phonological scores between initial- and post-learning phases.

Test	Initial-learning			Post-learning			r_c	$p(\text{unc})$	$p(\text{FDR})$	
	Mean	SD	Range	Mean	SD	Range				
Cognitive										
PS	75.5	9.5	(60.5, 91)	76.3	9.7	(61, 93)	0.26	0.313	0.452	ns
Inhibition	0.46	0.09	(0.27, 0.60)	0.46	0.06	(0.36, 0.57)	-0.10	0.729	0.729	ns
DigitSpan	9.1	1.6	(5.5, 11.5)	9.2	1.3	(7.5, 12.5)	0.19	0.571	0.710	ns
HKLLT	11.8	2.7	(7, 16)	12.8	2.6	(8, 16)	0.67	0.043	0.094	†
Raven	0.67	0.97	(-1.89, 2.05)	1.22	0.95	(-0.98, 2.40)	0.67	0.009	0.057	†
PicName	1358.7	221	(1026.4, 1821.9)	1247	276.2	(853.8, 1946.9)	-0.67	0.007	0.057	†
Fluency	12.6	2.6	(8.5, 17.3)	12.7	2.6	(8.3, 16.8)	0.14	0.601	0.710	ns
Phonological										
CanDiscrim	19.6	7.6	(7, 35)	22.3	6.7	(11, 39)	0.57	0.036	0.094	†
EngDiscrim	21	5.3	(8, 29)	23.2	5.9	(5, 29)	0.42	0.111	0.206	ns
ItaDiscrim	14.9	4.9	(6, 25)	16.9	5.4	(6, 26)	0.67	0.013	0.057	†
CanSpoon	4.1	6.4	(0, 23)	7.5	8.5	(0, 28)	0.62	0.031	0.094	†
EngSpoon	10.8	7.3	(0, 22)	11.8	7.4	(0, 24)	0.12	0.685	0.729	ns
ItaSpoon	5.2	5.5	(0, 19)	6	6.9	(0, 24)	0.32	0.224	0.364	ns

r_c , effect size of the Wilcoxon signed rank test, $p(\text{unc})$, uncorrected p , $p(\text{FDR})$, FDR-corrected p .

***, $p < 0.001$, **, $p < 0.01$, *, $p < 0.05$, †, marginal, ns, non-significant.

semantic retrieval strategies could have been utilized more in the Can-to-Ita task.

Besides HKLLT and semantic fluency, the remaining cognitive measures were kept in at most only one model, namely, in-class Ita-to-Can. In addition to processing speed, which is known to be a general deciding factor of the performance of older adults across domains (Salthouse, 1996), the in-class Ita-to-Can was associated with faster picture naming latency and better Raven's SPM. In contrast, in terms of phonological function measures, while EngSpoon contributed positively to in-class Can-to-Ita, the opposite was true for digit span forward (an index of phonological STM), suggesting that phonological function measures do not necessarily contribute positively. The inhibition function, measured using Stroop Color and Word Test, was also negatively to the learning performance.

Taken together, the anatomical and cognitive models of the learning performance were consistent with our main hypothesis that both semantic and episodic memory functions are prominently implicated in older adults. Phonological skills were also implicated, but their associations with learning performance were not as consistently observed.

4.2. Differential Hemispheric Associations in Immediate and Long-Term Retention

Beyond our main hypothesis, the present results could also be viewed from another perspective, in that there were interesting differences in the global pattern of anatomical associations between immediate learning and long-term retention (see **Figures 4, 5**). As discussed above, the left pars orbitalis and left caudal middle frontal cortex were the only two regions that consistently showed associations with the performance in both

immediate learning (Can-to-Ita) and long-term retention. All other regions were only associated with the performance of either the in-class Can-to-Ita or the final test score.

For the in-class Can-to-Ita performance, there were five such regions: left entorhinal cortex, right insula, right pars orbitalis, right caudal anterior cingulate, right middle temporal gyrus. Among them, significant anatomical associations with Can-to-Ita performance were found at left entorhinal cortex (see section 4.1) and right insula. The insula is known to be involved in procedural memory and rule-based learning (Ullman, 2001a, 2004; Yang and Li, 2012). In the present study, the in-class Can-to-Ita score was positively associated with the adjusted volume of the right insula but negatively associated with its adjusted surface area and curvature. In contrast to the right insula, the remaining three right hemisphere regions only showed some subtle changes in the association strength with the in-class Can-to-Ita score, in the sense that while the association was non-significant in any of the five weeks, the magnitude of such association varied significantly across weeks. Taken together, the significant association of right insula and the significant modulation of association over the other three right hemisphere regions suggest that the right hemisphere is implicated in the initial-learning phase. Such right hemisphere associations above added to the literature that reported some associations of the right hemisphere with FLL performance (Hosoda et al., 2013) or bilingualism-induced neuroplasticity (Bubbico et al., 2019; Maschio et al., 2019; Legault et al., 2019b), although the precise roles should be tested using online functional MRI tasks.

For long-term language retention, beyond left pars orbitalis and left caudal middle frontal cortex, seven regions were additionally associated with the final test results: left precentral gyrus (see section 4.1), left SMG (see section 4.1), left pars

triangularis, left rostral middle frontal cortex, left banks of superior temporal sulcus, right fusiform gyrus, and right entorhinal cortex. In other words, a more left-dominant network was implicated in long-term vocabulary retention, with seven left hemisphere regions showing significant associations with the final test score when compared to only two right hemisphere regions. Also, unlike the subtle right-hemispheric associations with in-class Can-to-Ita, all regions here showed significant associations with individual test scores. For the left hemisphere, the inclusion of left pars triangularis was not surprising, given that it is a part of the Broca's area. The left rostral middle frontal cortex has been implicated in language learning (Sheppard et al., 2012; Novén et al., 2019), while the left banks of superior temporal sulcus is known to reflect verbal intelligence and receptive language functions in children (Li et al., 2020). For the right hemisphere, the right fusiform gyrus is well-known for its role in object recognition (as opposed to word recognition for the left fusiform gyrus), while the right entorhinal cortex was especially important for long-term memory consolidation (Haist et al., 2001; Piefke et al., 2003).

Overall, while the best anatomical predictor of immediate success was found on semantics-associated (the left pars orbitalis) and episodic memory-related regions (left caudal middle frontal gyrus and the left entorhinal cortex), there were significant association with the right insula and some subtle changes in association with the right pars orbitalis, right rostral anterior cingulate, and right middle temporal gyrus over the course of learning. The increased left-lateralization in the association pattern of the final test scores was consistent with the view that the right-hemisphere homologs of the left-hemisphere language areas may provide "scaffolding" in the early phase of FLL, before a right-to-left shift occurs (Yang et al., 2015; Qi and Legault, 2020).

4.3. The Anatomical Associations of Sub-cortical Regions

In previous works, increased hippocampal volume was observed in Swedish individuals studying Italian (Bellander et al., 2016). The hippocampal volume was also a good predictor of the achieved vocabulary proficiency (Nilsson et al., 2021). Apart from the hippocampus, the striatum has been highlighted in previous works (Ullman, 2004; Abutalebi and Green, 2007; Li et al., 2014; Legault et al., 2019a). However, instead of the hippocampus and striatum, our results showed that the thalamus has stronger association with long-term vocabulary retention, with the cortical volumes of the bilateral thalami showing positive associations with the final test scores. The thalamus is implicated in language production by selecting lexical and semantic representations (Abutalebi and Green, 2016), and an expansion of thalamus is associated with greater second language immersion (Pliatsikas et al., 2017; Deluca et al., 2019a). The present finding added to the literature showing that the structural parameters of thalamus are sensitive to aspects of FLL, although the underlying reasons for the associations should be investigated further.

For hippocampus, we observed some degree of association of the right hippocampal volume with the Can-to-Ita test score,

but the association did not survive the FDR correction. Indeed, while the role of hippocampus in associative memory, especially in various learning and recall paradigms, is well-established in patient populations in the neurological literature (Suzuki, 2008), direct associations of hippocampal volume with associative memory have not been consistently found in cognitively normal older adults (Becker et al., 2015; Zheng et al., 2017). Finally, for striatum, the sub-cortical volumes were significantly associated with none of the three learning measures. One possibility is that these structures are non-unitary in functions. A finer-grained atlas may be necessary to reveal the associations.

4.4. Conclusion and Limitations

Previous studies have suggested that structural associations could provide valuable information for inferring functional associations (Kanai and Rees, 2011; Chiarello et al., 2013). In the present study, significant associations with vocabulary learning performances were consistently found for the left pars orbitalis and left caudal middle frontal cortex. These results suggest that the individual variations in structural morphometry of the prefrontal lobe are strongly associated with language learning success in older adults, considering that the prefrontal lobe is known to be involved in many "higher" cognitive skills such as language, reasoning, and planning (Wood and Grafman, 2003). In particular, the structural brain and cognitive models were consistent with our main hypothesis that semantic and episodic memory functions likely play an important role in language learning in older adults, with the regions most implicated with these functions being prominently represented in the overall pattern of associations. It should be noted that the present study had not explicitly tested what cognitive processes were involved during learning, given that an online design was not employed. Indeed, an alternative account for the association pattern observed is that the surface-based morphological features and the cognitive predictors are good indices of language aptitude, as previous studies suggested (e.g., Novén et al., 2019). Which could be a parsimonious and alternative explanation why these measures could significantly predict the FLL performance. Nonetheless, language aptitude also has multiple components that are differentially associated with cognitive functions, which could provide a more domain-general account of the anatomical associations with learning performance.

Our research complemented previous works that put a stronger focus on the acquisition of other linguistic elements such as phonology (Wong and Perrachione, 2007) and grammar (Yang and Li, 2012). However, due to the absence of a young learner group in the present study, no conclusion could be drawn regarding the relative contributions of semantic functions in FLL across older and younger adults. On the other hand, the majority of the studies thus far were on younger learners, and the role of the left pars orbitalis and left caudal middle frontal gyrus have seldom been reported. In this respect, the present findings suggest that semantic functions could play a more important role in older adults, although this hypothesis should be re-examined with online experimental designs.

Also, a strength of the present study was that objective measures of learning performance were collected throughout the Italian vocabulary learning protocol. Such longitudinal data play a crucial role in the sensitivity of the present analysis in revealing the anatomical markers of learning performance, despite our relatively modest sample size. In addition, by incorporating different surface-based measures into our univariate LME models, we successfully revealed a relatively consistent pattern of significant brain–learning associations. Where an association was found, the adjusted cortical volume was consistently a positive predictor of learning performance, with only one exception found at the fusiform gyrus. Similarly, the adjusted curvature of the regions involved was a strong positive predictor of learning performance, with only two exceptions found for the left entorhinal cortex and the right insula. In contrast, the adjusted area was consistently a negative predictor of the learning performance, without any exception. Adjusted thickness showed the least consistent pattern, being a positive predictor at four ROIs but a negative predictor at two ROIs.

The present work was limited in several ways. First, the MRI scans were not acquired at the beginning of the Italian learning programme but after an introductory Italian lesson of 90 min, to acclimate the learners with a more uniform experience with the target language prior to the scanning. Previous studies have shown that measurable changes in diffusion MRI indices could be induced by short-term intervention of 45 min (Tavor et al., 2020) or 120 min (Sagi et al., 2012). Structural changes were found after 20 days of language learning (Legault et al., 2019b), with some even estimating that only 2 h of training could bring about structural changes in the brain (Park and Bischof, 2013). However, considering that such structural changes may not be long-lasting, and the neuroplasticity of older adults' brains are lower than younger adults' (e.g., Freitas et al., 2011), it is unlikely that a 1.5-h lesson could significantly influence the anatomical measures extracted, to a point that their associations with learning performance would change qualitatively. Secondly, we acknowledge that from the experimenter's perspective, it would have been best to arrange the cognitive and phonological tests prior to the introductory Italian lesson. However, at the beginning of the programme, most participants were anxious and most motivated to start learning Italian. As such, arranging the prolonged testing of about 3.5 h that are not directly related to learning Italian could be seen as a non-ideal arrangement from the learner's perspective, which could lower their engagement in the programme, leading to an increased dropout rate. Because it was not our primary goal to investigate the learning-induced cognitive gains, a compromise was made to spread the cognitive and phonological tests over visits 3 to 6. Considering that a cognitive gain or an improvement of phonological skills could have already incurred following the 90-min introductory lesson and the 2-h vocabulary lessons in the initial-learning phase, we acknowledge that our decision to delay these assessments could have led to an underestimation of the cognition gains or phonological skill improvements reported.

Finally, no post-learning MRI scan was conducted, such that any learning-related neuroplasticity could not be tested. The lack of a younger group also precluded any formal hypothesis testing on the differential brain–learning associations between young and older learners. These issues will be addressed in our next round of data collection.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board, The Hong Kong Polytechnic University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MF, AA, and WW contributed to the conception and grant application of the study. MF and TL implemented the Italian learning protocol in the present study. MF, JC, TL, MM, and N-YH were involved in data collection and data transcription. MF and MM analyzed the data and wrote the manuscript. MF, MM, and WW contributed to the interpretation of the results. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partially supported by HKRGC-GRF grant 15606119 and Dean's Reserve from the Faculty of Humanities, The Hong Kong Polytechnic University, both awarded to WW. The article processing charge was supported by the Research Institute for Smart Aging (RISA), The Hong Kong Polytechnic University.

ACKNOWLEDGMENTS

The present analysis was conducted partially on the computer clusters offered by the University Facility of Big Data Analytics (UBDA) of the Hong Kong Polytechnic University. We thank two undergraduate student assistants (Kristy Wong and Sabrina Wong) for their assistance in data transcription. We also thank Joseph Tse of the MRI Unit, Department of Diagnostic Radiology, The University of Hong Kong, for scheduling and administrating our MRI sessions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2022.787413/full#supplementary-material>

REFERENCES

- Abutalebi, J., and Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguist.* 20, 242–275. doi: 10.1016/j.jneuroling.2006.10.003
- Abutalebi, J., and Green, D. W. (2016). Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Bilingualism Lang. Cogn.* 19, 689–698. doi: 10.1017/S1366728916000225
- Abutalebi, J., Guidi, L., Borsa, V., Canini, M., Rosa, P. A. D., Parris, B. A., et al. (2015). Bilingualism provides a neural reserve for aging populations. *Neuropsychologia* 69, 201–210. doi: 10.1016/j.neuropsychologia.2015.01.040
- Antonenko, D., Meinzer, M., Lindenberg, R., Witte, A. V., and Flöel, A. (2012). Grammar learning in older adults is linked to white matter microstructure and functional connectivity. *NeuroImage* 62, 1667–1674. doi: 10.1016/j.neuroimage.2012.05.074
- Antonou, M., Gunasekera, G. M., and Wong, P. C. M. (2013). Foreign language training as cognitive therapy for age-related cognitive decline: a hypothesis for future research. *Neurosci. Biobehav. Rev.* 37, 2689–2698. doi: 10.1016/j.neubiorev.2013.09.004
- Antonou, M., and Wright, S. M. (2017). Uncovering the mechanisms responsible for why language learning may promote healthy cognitive aging. *Front. Psychol.* 8, 2217. doi: 10.3389/fpsyg.2017.02217
- Baddeley, A., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev.* 105, 158–173. doi: 10.1037/0033-295X.105.1.158
- Bak, T. H., Long, M. R., Vega-Mendoza, M., and Sorace, A. (2016). Novelty, challenge, and practice: the impact of intensive language learning on attentional functions. *PLOS ONE* 11, e0153485. doi: 10.1371/JOURNAL.PONE.0153485
- Bakker-Marshall, I., Takashima, A., Fernandez, C. B., Janzen, G., McQueen, J. M., and Hell, J. G. V. (2021). Overlapping and distinct neural networks supporting novel word learning in bilinguals and monolinguals. *Bilingualism Lang. Cogn.* 24, 524–536. doi: 10.1017/S1366728920000589
- Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychonomic Bull. Rev.* 10, 344–380. doi: 10.3758/BF03196494
- Becker, N., Laukka, E. J., Kalpouzos, G., Naveh-Benjamin, M., Bäckman, L., and Brehmer, Y. (2015). Structural brain correlates of associative memory in older adults. *NeuroImage* 118, 146–153. doi: 10.1016/j.neuroimage.2015.06.002
- Bellander, M., Berggren, R., Mårtensson, J., Brehmer, Y., Wenger, E., Li, T. Q., et al. (2016). Behavioral correlates of changes in hippocampal gray matter structure during acquisition of foreign vocabulary. *NeuroImage* 131, 205–213. doi: 10.1016/j.neuroimage.2015.10.020
- Berggren, R., Nilsson, J., Brehmer, Y., Schmiedek, F., and Lövdén, M. (2020). Foreign language learning in older age does not improve memory or intelligence: Evidence from a randomized controlled study. *Psychol. Aging* 35, 212–219. doi: 10.1037/pag0000439
- Bialystok, E., Craik, F. I., and Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia* 45, 459–464. doi: 10.1016/j.neuropsychologia.2006.10.009
- Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/CERCOR/BHP055
- Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wolbers, T., et al. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage* 25, 958–968. doi: 10.1016/j.neuroimage.2004.12.019
- Bubbico, G., Chiacchiaretta, P., Parenti, M., di Marco, M., Panara, V., Sepede, G., et al. (2019). Effects of second language learning on the plastic aging brain: Functional connectivity, cognitive decline, and reorganization. *Front. Neurosci.* 13, 423. doi: 10.3389/fnins.2019.00423
- Chan, A., and Kwok, I. (1998). *Hong Kong List Learning Test (HKLLT): Manual and Preliminary Norms*. (Department of Psychology, The Chinese University of Hong Kong, Hong Kong).
- Chandrasekaran, B., Sampath, P. D., and Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *J. Acoust. Soc. America* 128, 456. doi: 10.1121/1.3445785
- Chen, L., Wu, J., Hartwigsen, G., Li, Z., Wang, P., and Feng, L. (2021). The role of a critical left fronto-temporal network with its right-hemispheric homologue in syntactic learning based on word category information. *J. Neurolinguist.* 58, 100977. doi: 10.1016/j.jneuroling.2020.100977
- Chiarello, C., Vazquez, D., Felton, A., and Leonard, C. M. (2013). Structural asymmetry of anterior insula: behavioral correlates and individual differences. *Brain Lang.* 126, 109–122. doi: 10.1016/j.bandl.2013.03.005
- Chung, Y. S., Hyatt, C. J., and Stevens, M. C. (2017). Adolescent maturation of the relationship between cortical gyrification and cognitive ability. *NeuroImage* 158, 319–331. doi: 10.1016/j.neuroimage.2017.06.082
- Costa, A., and Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nat. Rev. Neurosci.* 15, 336–345. doi: 10.1038/nrn3709
- Daselaar, S. M., Veltman, D. J., and Witter, M. P. (2004). Common pathway in the medial temporal lobe for storage and recovery of words as revealed by event-related functional MRI. *Hippocampus* 14, 163–169. doi: 10.1002/HIPO.10158
- DeLuca, V., Rothman, J., Bialystok, E., and Pliatsikas, C. (2019a). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proc. Natl. Acad. Sci.* 116, 7565–7574. doi: 10.1073/PNAS.1811513116
- Deluca, V., Rothman, J., and Pliatsikas, C. (2019b). Linguistic immersion and structural effects on the bilingual brain: a longitudinal study. *Bilingualism Lang. Cogn.* 22, 1160–1175. doi: 10.1017/S1366728918000883
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism Lang. Cogn.* 18, 145–162. doi: 10.1017/S1366728913000722
- Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci.* 1, 41–50. doi: 10.1038/35036213
- Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., and Lipton, P. (2012). Towards a functional organization of episodic memory in the medial temporal lobe. *Neurosci. Biobehav. Rev.* 36, 1597–1608. doi: 10.1016/j.neubiorev.2011.07.006
- Fischl, B. (2012). Freesurfer. *NeuroImage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021
- Fischl, B., and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci.* 97, 11050–11055. doi: 10.1073/PNAS.200033797
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355. doi: 10.1016/S0896-6273(02)00569-X
- Fong, M. C.-M., Hui, N. Y., Fung, E. S.-W., Ma, M. K.-H., Law, T. S.-T., Wang, X., et al. (2020). Which cognitive functions subserve clustering and switching in category fluency? generalisations from an extended set of semantic categories using linear mixed-effects modelling. *Quart. J. Exp. Psychol.* 73, 2132–2147. doi: 10.1177/1747021820957135
- Fong, M. C.-M., Law, T. S. T., Ma, M. K.-H., Hui, N. Y., and Wang, W. S. (2021). Can inhibition deficit hypothesis account for age-related differences in semantic fluency? Converging evidence from Stroop color and word test and an ERP flanker task. *Brain Lang.* 218, 104952. doi: 10.1016/j.bandl.2021.104952
- Ford, A. A., Triplett, W., Sudhyadhom, A., Gullett, J., McGregor, K., FitzGerald, D. B., et al. (2013). Broca's area and its striatal and thalamic connections: A diffusion-MRI tractography study. *Front. Neuroanatomy* 7, 8. doi: 10.3389/FNANA.2013.00008
- Freitas, C., Perez, J., Knobel, M., Tormos, J. M., Oberman, L., Eldaief, M., et al. (2011). Changes in cortical plasticity across the lifespan. *Front. Aging Neurosci.* 3, 5. doi: 10.3389/FNAGI.2011.00005
- Fuhrmeister, P., and Myers, E. B. (2021). Structural variation in the temporal lobe predicts learning and retention of non-native speech sounds. *Lang. Cogni. Neurosci.* 37, 63–79. doi: 10.1080/23273798.2021.1944658
- Gautam, P., Anstey, K. J., Wen, W., Sachdev, P. S., and Cherbuin, N. (2015). Cortical gyrification and its relationships with cortical volume, cortical thickness, and cognitive performance in healthy mid-life adults. *Behav. Brain Res.* 287, 331–339. doi: 10.1016/j.bbr.2015.03.018

- Gillon, G. T. (2017). *Phonological Awareness: From Research To Practice, 2nd edn.* (New York: Guilford Publications).
- Glass, B. D., Chotibut, T., Pacheco, J., Schnyer, D. M., and Maddox, W. T. (2012). Normal aging and the dissociable prototype learning systems. *Psychol. Aging* 27, 120–128. doi: 10.1037/A0024971
- Golden, C., and Freshwater, S. (1978). *Stroop Color and Word Test.* Chicago, IL: Stoelting.
- Golestani, N., and Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain Lang.* 109, 55–67. doi: 10.1016/J.BANDL.2008.01.005
- Grant, A., Dennis, N. A., and Li, P. (2014). Cognitive control, cognitive reserve, and memory in the aging bilingual brain. *Front. Psychol.* 5, 1401. doi: 10.3389/FPSYG.2014.01401
- Grasby, K. L., Jahanshad, N., Painter, J. N., Colodro-Conde, L., Bralten, J., Hibar, D. P., et al. (2020). The genetic architecture of the human cerebral cortex. *Science* 367, 6484. doi: 10.1126/science.aay6690
- Green, S., Blackmon, K., Thesen, T., DuBois, J., Wang, X., Halgren, E., et al. (2018). Parieto-frontal gyrification and working memory in healthy adults. *Brain Imag. Behav.* 12, 303–308. doi: 10.1007/S11682-017-9696-9
- Grogan, A., Green, D. W., Ali, N., Crinion, J. T., and Price, C. J. (2009). Structural correlates of semantic and phonemic fluency ability in first and second languages. *Cereb. Cortex* 19, 2690–2698. doi: 10.1093/CERCOR/BHP023
- Haist, F., Gore, J. B., and Mao, H. (2001). Consolidation of human memory over decades revealed by functional magnetic resonance imaging. *Nat. Neurosci.* 4, 1139–1145. doi: 10.1038/nn739
- Hämäläinen, S., Joutsa, J., Sihvonen, A. J., Leminen, A., and Lehtonen, M. (2018). Beyond volume: A surface-based approach to bilingualism-induced grey matter changes. *Neuropsychologia* 117, 1–7. doi: 10.1016/J.NEUROPSYCHOLOGIA.2018.04.038
- Hartshorne, J. K., and Germine, L. T. (2015). When does cognitive functioning peak? the asynchronous rise and fall of different cognitive abilities across the life span. *Psychol. Sci.* 26, 433–443. doi: 10.1177/0956797614567339
- Hartshorne, J. K., Tenenbaum, J. B., and Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million english speakers. *Cognition* 177, 263–277. doi: 10.1016/J.COGNITION.2018.04.007
- Harvey, D. Y., Wei, T., Ellmore, T. M., Hamilton, A. C., and Schnur, T. T. (2013). Neuropsychological evidence for the functional role of the uncinate fasciculus in semantic control. *Neuropsychologia* 51, 789–801. doi: 10.1016/J.NEUROPSYCHOLOGIA.2013.01.028
- Hasher, L., Stoltzfus, E. R., Zacks, R. T., and Rypma, B. (1991). Age and inhibition. *J. Exp. Psychol. Learn. Memory Cogn.* 17, 163–169. doi: 10.1037//0278-7393.17.1.163
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Hogstrom, L. J., Westlye, L. T., Walhovd, K. B., and Fjell, A. M. (2013). The structure of the cerebral cortex across adult life: Age-related patterns of surface area, thickness, and gyrification. *Cereb. Cortex* 23, 2521–2530. doi: 10.1093/CERCOR/BHS231
- Hosoda, C., Tanaka, K., Nariyai, T., Honda, M., and Hanakawa, T. (2013). Dynamic neural network reorganization associated with second language vocabulary acquisition: a multimodal imaging study. *J. Neurosci.* 33, 13663–13672. doi: 10.1523/JNEUROSCI.0410-13.2013
- Husa, R. A., Gordon, B. A., Cochran, M. M., Bertolin, M., Bond, D. N., and Kirchoff, B. A. (2017). Left caudal middle frontal gray matter volume mediates the effect of age on self-initiated elaborative encoding strategies. *Neuropsychologia* 106, 341–349. doi: 10.1016/j.neuropsychologia.2017.10.004
- Ingalvalson, E. M., Nowicki, C., Zong, A., and Wong, P. C. M. (2017). Non-native speech learning in older adults. *Front. Psychol.* 8, 148. doi: 10.3389/FPSYG.2017.00148
- Kanai, R., and Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nat. Rev. Neurosci.* 12, 231–242. doi: 10.1038/nrn3000
- Kepinska, O., de Rover, M., Caspers, J., and Schiller, N. O. (2017). On neural correlates of individual differences in novel grammar learning: an fMRI study. *Neuropsychologia* 98, 156–168. doi: 10.1016/J.NEUROPSYCHOLOGIA.2016.06.014
- Kepinska, O., de Rover, M., Caspers, J., and Schiller, N. O. (2018). Connectivity of the hippocampus and Broca's area during acquisition of a novel grammar. *NeuroImage* 165, 1–10. doi: 10.1016/J.NEUROIMAGE.2017.09.058
- Khoi, R., and Shariffar, S. (2013). Memorization versus semantic mapping in L2 vocabulary acquisition. *ELT J.* 67, 199–209. doi: 10.1093/ELT/CCS101
- King, B. M., Rosopa, P., and Minium, E. W. (2011). *Statistical Reasoning in the Behavioral Sciences, 6th Edn.*, vol. 394 (Hoboken, NJ: J. Wiley & Sons).
- Kirchoff, B. A., Gordon, B. A., and Head, D. (2014). Prefrontal gray matter volume mediates age effects on memory strategies. *NeuroImage* 90, 326–334. doi: 10.1016/J.NEUROIMAGE.2013.12.052
- Klein, D., Mok, K., Chen, J. K., and Watkins, K. E. (2014). Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain Lang.* 131, 20–24. doi: 10.1016/J.BANDL.2013.05.014
- Klimova, B. (2018). Learning a foreign language: A review on recent findings about its effect on the enhancement of cognitive functions among healthy older individuals. *Front. Hum. Neurosci.* 12, 305. doi: 10.3389/FNHUM.2018.00305
- Klimova, B., and Pikhart, M. (2020). Current research on the impact of foreign language learning among healthy seniors on their cognitive functions from a positive psychology perspective—a systematic review. *Front. Psychol.* 11, 765. doi: 10.3389/FPSYG.2020.00765
- Koda, K. (1998). The role of phonemic awareness in second language reading. *Second Lang. Res.* 14, 194–215. doi: 10.1191/026765898676398460
- Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the n400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532
- Lee, H., Devlin, J. T., Shakeshaft, C., Stewart, L. H., Brennan, A., Glensman, J., et al. (2007). Anatomical traces of vocabulary acquisition in the adolescent brain. *J. Neurosci.* 27, 1184–1189. doi: 10.1523/JNEUROSCI.4442-06.2007
- Legault, J., Fang, S. Y., Lan, Y. J., and Li, P. (2019a). Structural brain changes as a function of second language vocabulary training: Effects of learning context. *Brain Cogn.* 134, 90–102. doi: 10.1016/J.BANDC.2018.09.004
- Legault, J., Grant, A., Fang, S. Y., and Li, P. (2019b). A longitudinal investigation of structural brain changes during second language learning. *Brain Lang.* 197, 104661. doi: 10.1016/J.BANDL.2019.104661
- Lenneberg, E. H. (1967). *The Biological Foundations of Language* (New York: Wiley). doi: 10.1080/21548331.1967.11707799
- Li, P., and Grant, A. (2016). Second language learning success revealed by brain networks. *Bilingualism Lang. Cogn.* 19, 657–664. doi: 10.1017/S1366728915000280
- Li, P., Legault, J., and Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. *Cortex* 58, 301–324. doi: 10.1016/J.CORTEX.2014.05.001
- Li, T., McCorkle, G. S., Williams, D. K., Badger, T. M., and Ou, X. (2020). Cortical morphometry is associated with neuropsychological function in healthy 8-year-old children. *J. Neuroimag.* 30, 833–842. doi: 10.1111/JON.12754
- Linck, J. A., Osthus, P., Koeth, J. T., and Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bull. Rev.* 21, 861–883. doi: 10.3758/S13423-013-0565-2
- Luders, E., Kurth, F., Mayer, E. A., Toga, A. W., Narr, K. L., and Gaser, C. (2012). The unique brain anatomy of meditation practitioners: alterations in cortical gyrification. *Front. Hum. Neurosci.* 6, 34. doi: 10.3389/FNHUM.2012.00034
- Luk, G., Bialystok, E., Craik, F. I. M., and Grady, C. L. (2011). Lifelong bilingualism maintains white matter integrity in older adults. *J. Neurosci.* 31, 16808–16813. doi: 10.1523/JNEUROSCI.4563-11.2011
- Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., et al. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage* 63, 240–244. doi: 10.1016/J.NEUROIMAGE.2012.06.043
- Mackey, A., and Sachs, R. (2012). Older learners in SLA research: a first look at working memory, feedback, and L2 development. *Lang. Learn.* 62, 704–740. doi: 10.1111/J.1467-9922.2011.00649.X
- Maschio, N. D., Fedeli, D., Sulpizio, S., and Abutalebi, J. (2019). The relationship between bilingual experience and gyrification in adulthood: A cross-sectional surface-based morphometry study. *Brain Lang.* 198, 104680. doi: 10.1016/J.BANDL.2019.104680
- McClelland, J. L., McNaughton, B. L., and O'reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights

- from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457. doi: 10.1037/0033-295X.102.3.419
- Mestres-Missé, A., Càmará, E., Rodríguez-Fornells, A., Rotte, M., and Münte, T. F. (2008). Functional neuroanatomy of meaning acquisition from context. *J. Cogn. Neurosci.* 20, 2153–2166. doi: 10.1162/JOCN.2008.20150
- Nilsson, J., Berggren, R., Garzón, B., Lebedev, A. V., and Lövdén, M. (2021). Second language learning in older adults: effects on brain structure and predictors of learning success. *Front. Aging Neurosci.* 13, 666851. doi: 10.3389/FNAGI.2021.666851
- Novén, M., Schremm, A., Nilsson, M., Horne, M., and Roll, M. (2019). Cortical thickness of broca's area and right homologue is related to grammar learning aptitude and pitch discrimination proficiency. *Brain Lang.* 188, 42–47. doi: 10.1016/J.BANDL.2018.12.002
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Olsen, R. K., Pangelinan, M. M., Bogulski, C., Chakravarty, M. M., Luk, G., Grady, C. L., et al. (2015). The effect of lifelong bilingualism on regional grey and white matter volume. *Brain Res.* 1612, 128–139. doi: 10.1016/J.BRAINRES.2015.02.034
- Paap, K. R., Johnson, H. A., and Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex* 69, 265–278. doi: 10.1016/J.CORTEX.2015.04.014
- Papagno, C., Valentine, T., and Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *J. Mem. Lang.* 30, 331–347. doi: 10.1016/0749-596X(91)90040-Q
- Park, D. C., and Bischof, G. N. (2013). The aging mind: neuroplasticity in response to cognitive training. *Dialogues Clin. Neurosci.* 15, 109. doi: 10.31887/DCNS.2013.15.1/DPARK
- Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? the representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987. doi: 10.1038/nrn2277
- Petrides, M., Alivisatos, B., Meyer, E., and Evans, A. C. (1993). Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proc. Natl. Acad. Sci.* 90, 878–882. doi: 10.1073/PNAS.90.3.878
- Pfenninger, S. E., and Polz, S. (2018). Foreign language learning in the third age: A pilot feasibility study on cognitive, socio-affective and linguistic drivers and benefits in relation to previous bilingualism of the learner. *J. Eur. Second Lang. Assoc.* 2, 1. doi: 10.22599/JESLA.36
- Piefke, M., Weiss, P. H., Zilles, K., Markowitsch, H. J., and Fink, G. R. (2003). Differential remoteness and emotional tone modulate the neural correlates of autobiographical memory. *Brain* 126, 650–668. doi: 10.1093/BRAIN/AWG064
- Pliatsikas, C., DeLuca, V., Moschopoulou, E., and Saddy, J. D. (2017). Immersive bilingualism reshapes the core of the brain. *Brain Struct. Funct.* 222, 1785–1795. doi: 10.1007/S00429-016-1307-9
- Qi, Z., Han, M., Wang, Y., de los Angeles, C., Liu, Q., Garel, K., et al. (2019). Speech processing and plasticity in the right hemisphere predict variation in adult foreign language learning. *NeuroImage* 192, 76–87. doi: 10.1016/J.NEUROIMAGE.2019.03.008
- Qi, Z., and Legault, J. (2020). Neural hemispheric organization in successful adult language learning: is left always right? *Psychol. Learn. Motivation Adv. Res. Theory* 72, 119–163. doi: 10.1016/BS.PLM.2020.02.004
- Ralph, M. A. L., Jefferies, E., Patterson, K., and Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18, 42–55. doi: 10.1038/nrn.2016.150
- Raven, J. C., and Court, J. H. (1998). *Raven's Progressive Matrices and Vocabulary Scales*, vol. 759. (Oxford, UK: Oxford Psychologists Press).
- Sabb, F. W., Bilder, R. M., Chou, M., and Bookheimer, S. Y. (2007). Working memory effects on semantic processing: priming differences in pars orbitalis. *NeuroImage* 37, 311–322. doi: 10.1016/J.NEUROIMAGE.2007.04.050
- Sagi, Y., Tavor, I., Hofstetter, S., Tzur-Moryosef, S., Blumenfeld-Katzir, T., and Assaf, Y. (2012). Learning in the fast lane: New insights into neuroplasticity. *Neuron* 73, 1195–1203. doi: 10.1016/J.NEURON.2012.01.025
- Salthouse, T. (2012). Consequences of age-related cognitive declines. *Ann. Rev. Psychol.* 63, 201–226. doi: 10.1146/ANNUREV-PSYCH-120710-100328
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol. Rev.* 103, 403–428. doi: 10.1037/0033-295X.103.3.403
- Schroeder, S. R., and Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *J. Cogn. Psychol.* 24, 591–601. doi: 10.1080/20445911.2012.669367
- Service, E., and Craik, F. I. (1993). Differences between young and older adults in learning a foreign vocabulary. *J. Memory Lang.* 32, 608–623. doi: 10.1006/JMLA.1993.1031
- Shen, J., Wright, R., and Souza, P. E. (2016). On older listeners' ability to perceive dynamic pitch. *J. Speech Lang. Hearing Res.* 59, 572. doi: 10.1044/2015_JSLHR-H-15-0228
- Sheppard, J. P., Wang, J. P., and Wong, P. C. (2012). Large-scale cortical network properties predict future sound-to-word learning success. *J. Cogn. Neurosci.* 24, 1087. doi: 10.1162/JOCN_A_00210
- Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., and Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *J. Phonetics* 50, 99–119. doi: 10.1016/J.WOCN.2015.03.001
- Sliwiska, M. W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F., and Devlin, J. T. (2012). Early and sustained supramarginal gyrus contributions to phonological processing. *Front. Psychol.* 3, 161. doi: 10.3389/FPSYG.2012.00161
- Stein, M., Federspiel, A., Koenig, T., Wirth, M., Strik, W., Wiest, R., et al. (2012). Structural plasticity in the language system related to increased second language proficiency. *Cortex* 48, 458–465. doi: 10.1016/J.CORTEX.2010.10.007
- Suzuki, W. A. (2008). Associative learning signals in the brain. *Progr. Brain Res.* 169, 305–320. doi: 10.1016/S0079-6123(07)00019-2
- Tadayon, E., Pascual-Leone, A., and Santarnecchi, E. (2020). Differential contribution of cortical thickness, surface area, and gyrification to fluid and crystallized intelligence. *Cereb. Cortex* 30, 215–225. doi: 10.1093/CERCOR/BHZ082
- Tagarelli, K. M., Shattuck, K. F., Turkeltaub, P. E., and Ullman, M. T. (2019). Language learning in the adult brain: a neuroanatomical meta-analysis of lexical and grammatical learning. *NeuroImage* 193, 178–200. doi: 10.1016/J.NEUROIMAGE.2019.02.061
- Tavor, I., Botvinik-Nezer, R., Bernstein-Eliav, M., Tsarfay, G., and Assaf, Y. (2020). Short-term plasticity following motor sequence learning revealed by diffusion magnetic resonance imaging. *Hum. Brain Map.* 41, 442–452. doi: 10.1002/HBM.24814
- Thomas, M. H., and Wang, A. Y. (1996). Learning by the keyword mnemonic: Looking for long-term benefits. *J. Exp. Psychol. Appl.* 2, 330–342. doi: 10.1037/1076-898X.2.4.330
- Turk, D. J., Gillespie-Smith, K., Krigolson, O. E., Havard, C., Conway, M. A., and Cunningham, S. J. (2015). Selfish learning: the impact of self-referential encoding on children's literacy attainment. *Learn. Inst.* 40, 54–60. doi: 10.1016/J.LEARNINSTRUC.2015.08.001
- Ullman, M. T. (2001a). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism Lang. Cogn.* 4, 105–122. doi: 10.1017/S1366728901000220
- Ullman, M. T. (2001b). A neurocognitive perspective on language: the declarative/procedural model. *Nat. Rev. Neurosci.* 2, 717–726. doi: 10.1038/35094573
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231–270. doi: 10.1016/J.COGNITION.2003.10.008
- Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In eds G. Hickok, and S. L. Small, *Neurobiol. Lang.* Elsevier, p. 953–968. doi: 10.1016/B978-0-12-407794-2.00076-6
- Valis, M., Slaninova, G., Prazak, P., Poulouva, P., Kacetl, J., and Klimova, B. (2019). Impact of learning a foreign language on the enhancement of cognitive functions among healthy older population. *J. Psycholinguist. Res.* 48, 1311–1318. doi: 10.1007/S10936-019-09659-6
- van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., et al. (2019). Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? a systematic review. *Behav. Sci.* 9, 27. doi: 10.3390/BS9030027

- Verissimo, J., Verhaeghen, P., Goldman, N., Weinstein, M., and Ullman, M. T. (2021). Evidence that ageing yields improvements as well as declines across attention and executive functions. *Nat. Hum. Behav.* 6, 97–110. doi: 10.1038/s41562-021-01169-7
- Veroude, K., Norris, D. G., Shumskaya, E., Gullberg, M., and Indefrey, P. (2010). Functional connectivity between brain regions involved in learning words of a new language. *Brain Lang.* 113, 21–27. doi: 10.1016/J.BANDL.2009.12.005
- Wang, W. S.-Y. (2018). Critical periods for language: Comment on “Rethinking foundations of language from a multidisciplinary perspective” by T. Gong et al. *Phys. Life Rev.* 26–27, 179–183. doi: 10.1016/J.PLREV.2018.09.002
- Ware, C., Damnee, S., Djabelkhir, L., Cristancho, V., Wu, Y.-H., Benovici, J., et al. (2017). Maintaining cognitive functioning in healthy seniors with a technology-based foreign language program: a pilot feasibility study. *Front. Aging Neurosci.* 9, 42. doi: 10.3389/FNAGI.2017.00042
- Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., et al. (2010). Cortical thickness or grey matter volume? the importance of selecting the phenotype for imaging genetics studies. *NeuroImage* 53, 1135–1146. doi: 10.1016/J.NEUROIMAGE.2009.12.028
- Wong, A., Xiong, Y. Y., Kwan, P. W., Chan, A. Y., Lam, W. W., Wang, K., et al. (2009). The validity, reliability and clinical utility of the hong kong montreal cognitive assessment (HK-MoCA) in patients with cerebral small vessel disease. *Dementia Geriatric Cogn. Dis.* 28, 81–87. doi: 10.1159/000232589
- Wong, P. C. M., Ou, J., Pang, C. W. Y., Zhang, L., Tse, C. S., Lam, L. C. W., et al. (2019). Language training leads to global cognitive improvement in older adults: a preliminary study. *J. Speech Lang. Hearing* 62, 2411–2424. doi: 10.1044/2019_JSLHR-L-18-0321
- Wong, P. C. M., and Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native english-speaking adults. *Appl. Psycholinguist.* 28, 565–585. doi: 10.1017/S0142716407070312
- Wong, P. C. M., Perrachione, T. K., and Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Hum. Brain Map.* 28, 995–1006. doi: 10.1002/HBM.20330
- Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., et al. (2008). Volume of left heschl’s gyrus and linguistic pitch learning. *Cereb. Cortex* 18, 828–836. doi: 10.1093/CERCOR/BHM115
- Wood, J. N., and Grafman, J. (2003). Human prefrontal cortex: processing and representational perspectives. *Nat. Rev. Neurosci.* 4, 139–147. doi: 10.1038/nrn1033
- Yang, D. Y., Beam, D., Pelphrey, K. A., Abdullahi, S., and Jou, R. J. (2016). Cortical morphological markers in children with autism: a structural magnetic resonance imaging study of thickness, area, volume, and gyrification. *Mol. Autism* 7, 11. doi: 10.1186/S13229-016-0076-X
- Yang, J., Gates, K. M., Molenaar, P., and Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *J. Neurolinguist.* 33, 29–49. doi: 10.1016/J.JNEUROLING.2014.09.004
- Yang, J., and Li, P. (2012). Brain networks of explicit and implicit learning. *PLOS ONE* 7, e42993. doi: 10.1371/JOURNAL.PONE.0042993
- Yang, W., Liu, P., Cui, Q., Wei, D., Li, W., Qiu, J., et al. (2013). Directed forgetting of negative self-referential information is difficult: an fMRI study. *PLOS ONE* 8, e75190. doi: 10.1371/JOURNAL.PONE.0075190
- Yu, Q., McCall, D. M., Homayouni, R., Tang, L., Chen, Z., Schoff, D., et al. (2018). Age-associated increase in mnemonic strategy use is linked to prefrontal cortex development. *NeuroImage* 181, 162–169. doi: 10.1016/J.NEUROIMAGE.2018.07.008
- Yum, Y. N., Midgley, K. J., Holcomb, P. J., and Grainger, J. (2014). An ERP study on initial second language vocabulary learning. *Psychophysiology* 51, 364–373. doi: 10.1111/PSYP.12183
- Zatorre, R. J. (2013). Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science* 342, 585–589. doi: 10.1126/SCIENCE.1238414
- Zheng, Z., Li, R., Xiao, F., He, R., Zhang, S., and Li, J. (2017). Sex matters: hippocampal volume predicts individual differences in associative memory in cognitively normal older women but not men. *Front. Hum. Neurosci.* 11, 93. doi: 10.3389/FNHUM.2017.00093

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Fong, Ma, Chui, Law, Hui, Au and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.