

Title: Beyond volume: A surface-based approach to bilingualism-induced grey matter changes

Short title: Bilingualism-induced surface-based grey matter changes

Authors: Sini Hämäläinen^{1,2}, Juho Joutsa³⁻⁵, Aleks J. Sihvonen^{1,6}, Alina Leminen¹, Minna Lehtonen^{1,7,8}

¹ Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, 00014, Helsinki, Finland

² Advanced Magnetic Imaging Centre, Aalto NeuroImaging, Aalto University Espoo, Finland

³ Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA

⁴ Berenson-Allen Center for Noninvasive Brain Stimulation, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA

⁵ Department of Neurology, University of Turku, Turku, Finland

⁶ Faculty of Medicine, University of Turku, Finland

⁷ Department of Psychology, Abo Akademi University, Turku, Finland

⁸ MultiLing Center for Multilingualism in Society across the Lifespan, Department of Linguistics and Scandinavian studies, University of Oslo, Oslo, Norway.

Address all correspondence to:

Sini Hämäläinen, Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, Siltavuorenpenger 1 B, 00014 University of Helsinki, Helsinki, Finland.

Email: sini.hamalainen@helsinki.fi, Phone: +358 44 022 0030

Keywords: bilingualism, SBA, surface area, thickness, age of L2 acquisition

Abstract

Bilingualism is a sustained experience associated with structural changes in cortical grey matter (GM) morphology. Apart from a few studies, a dominant method used to assess bilingualism-induced GM changes has been the voxel-based morphometry (VBM) analysis. While VBM is sensitive to GM volume/density differences in general, it cannot be used to identify whether the observed difference is due to relative changes in, e.g., cortical thickness, area or folding, as it uses a single combined measure of them all. Here, we used surface-based analysis (SBA) approach to investigate whether early acquisition of a second language (L2) affects the cortical GM morphology relative to late L2 acquisition. More specifically, our aim was to test a hypothesis that early acquisition of two languages induces GM changes that are predominantly surface area-driven, while late acquisition is supposedly characterised with primarily thickness-driven changes. To this end, several surface-based measures were concurrently compared between the groups. In line with the hypothesis, the results revealed that early bilingual experience is associated with significantly extended cortical surface area over the left pars opercularis and the right superior temporal gyrus. Contrary to our expectations, however, we found no evidence supporting the postulated association between late L2 acquisition and increased cortical thickness. Nevertheless, our study highlights the importance of including cortical surface measures when investigating bilingualism-related GM modulations.

1 Introduction

Recently, neurocognitive research on bilingualism has increasingly focused attention on structural brain plasticity related to second language (L2) acquisition and use. Pioneering studies have investigated whether bilingualism is related to quantifiable changes in both cortical grey matter (GM) morphology as well as in underlying white matter (WM) fibre arrangement (for reviews, see e.g., Li et al. 2014; Stein et al. 2014; García-Pentón et al. 2015). While an emerging consensus posits that bilingualism is indeed characterized by both GM and WM structural modulations, the experimental evidence remains relatively scarce and highly varying, regarding, in particular, the GM modulations.

To date, most studies targeting bilingualism-induced GM changes have utilised a voxel-based morphometry (VBM) analysis, which allows for a whole-brain level voxel-by-voxel comparison of the GM density/volume between two groups (Ashburner and Friston, 2000). With VBM, bilingualism-induced increases in GM volume/density have been observed in several fronto-temporal and medial language-related areas (Abutalebi et al., 2015b, 2014; Kaiser et al., 2015; Mechelli et al., 2004) as well as in the cerebellum (Pliatsikas et al., 2014). In contrast to these findings, several other studies using the same whole-brain level approach have not been able to detect any global GM differences associated with bilingualism (Gold et al., 2013; Grogan et al., 2012; Ressel et al., 2012). In two of these studies, however, significant group differences appeared when the number of voxelwise comparisons was constrained to specific, a priori chosen structures; these effects showed increased GM volume bilaterally in Heschl's gyrus (Ressel et al., 2012) and in the right posterior supramarginal gyrus (Grogan et al., 2012) in relation to bilingual/multilingual exposure. Other studies using a constrained VBM analysis and/or a mean volume ROI analysis with varying samples have reported bilingualism-induced increases in GM volume also in bilateral temporal poles and

orbito-frontal cortices (Abutalebi et al., 2014), bilateral inferior parietal lobules and in the left putamen (Abutalebi et al., 2015a). Although these constrained analyses corroborate the whole-brain level findings in showing that bilingualism can induce structural GM modulations (but see Olsen et al., 2015, with no significant group effects despite a ROI approach), the reported effects have been scattered across the brain, showing little or no consistency between studies in general.

In addition to VBM studies and mean volume analyses, structural GM modulations can also be investigated by using a surface-based analysis (SBA) approach. SBA approach defines the cortical surface as a geometrical mesh model, created by extracting the inner and outer GM boundaries. Several cortical measures such as thickness, surface area, volume and mean curvature can be estimated from the mesh model by calculating distances between the vertices. While SBA has been regularly used to investigate, for instance, changes in cortical volume and thickness associated with typical and atypical development (Wierenga et al. 2014; Zielinski et al. 2014; Wallace et al. 2015), only two prior studies have implemented it to investigate bilingualism-induced GM modulations. Klein and colleagues (2014) performed a cross-sectional study on cortical thickness between monolinguals vs. simultaneous (age of acquisition of L2, AoA = < 3 years), early sequential (AoA = 4–7 years) and late sequential bilinguals (AoA = 8–13 years) and observed differences in the left and right IFG. In the left IFG, the early and late sequential bilinguals had significantly increased cortical thickness compared to monolinguals and simultaneous bilinguals. In the right IFG, however, the effect was reversed, and the difference was significant only between the late bilinguals and monolinguals (i.e., late bilinguals had thinner cortex than monolinguals). In another study, cortical thickness was investigated in participants who underwent intensive language training during three months (Mårtensson et al., 2012). They showed that gaining proficiency in the target language was associated with increased cortical thickness in the left fronto-temporal cortices. More specifically, changes were observed in the left IFG, the middle frontal gyrus and in the superior temporal gyrus

(STG); however, only the changes in the STG were significantly correlated with an increase in the target language proficiency. Notably, these SBA results are somewhat more concordant than the VBM/ROI findings in indicating, for example, the left IFG as a potential site for bilingualism-induced GM changes.

Considering that GM volume is the product of surface area and thickness (Panizzon et al., 2009), a major shortcoming of the previous studies examining bilingualism-related changes in the cortical GM morphology, is that these measures have not been investigated concurrently. First, accumulating evidence from several studies on cortical development and cellular architecture suggests that surface area and thickness are independent rather than redundant features of cortical structure (Im et al., 2008; Panizzon et al., 2009) with little to no covariation between them (Pakkenberg and Gundersen, 1997). Consequently, comparisons restricted to the GM volume alone risk obstructing bilingualism-related modulations that might be specific to either surface area or cortical thickness. In other words, two groups could, in theory, exhibit identical GM volume and still differ significantly from each other with respect to the underlying surface area to thickness ratio (for illustration, see Fig. 1). Conversely, two studies conducted with different bilingual samples could reveal seemingly concordant GM volume findings, while, in fact, the underlying cause for the observed GM difference (surface area vs. thickness-driven) might differ substantially between the studies.

Second, cortical surface area and thickness have been suggested to reflect distinct cellular mechanisms (Rakic, 2009; Raznahan et al., 2011), inasmuch as cortical surface area is thought to reflect the number of neuronal columns perpendicular to the pial surface, whereas cortical thickness relates to the number (Rakic, 2009) and, possibly, to the size of cells within these columns (Panizzon et al., 2009). Crucially, cortical surface area has been shown to increase rapidly through maturation

during infancy (0-2 years of age; Lyall et al., 2015) and early childhood (1-6 years of age; Remer et al., 2017), and to remain relatively stable from thereafter (Tamnes et al., 2017). Cortical thickness, on the other hand, appears well established at birth (Lyall et al., 2015), but has been suggested to alter dynamically across the entire life span in response to experience-dependent learning (Meyer et al., 2014). As the brain is considered to be particularly susceptible to the effects of the environment during periods of rapid growth (Lyall et al., 2015), early acquisition of two languages could be hypothesised to induce GM changes that are predominantly surface area-driven, whereas acquiring an L2 later in life, after surface area development has already been largely set, might lead to primarily thickness-driven GM changes. To test this hypothesis, we used SBA approach to investigate whether early exposure to L2 affects cortical morphology relative to late L2 exposure. Comparisons of several cortical measures on both whole-brain level as well as within a priori defined ROIs relevant for language processing were conducted. The ROI analysis was motivated by the reviewed literature suggesting that some of the effects might be discoverable only by limiting the analysis to specific language-related regions (Grogan et al., 2012; Ressel et al., 2012). We therefore included five bilateral ROIs encompassing areas where structural effects related to bilingualism or language processing have typically appeared; the inferior parietal cortex (IPC; Mechelli et al., 2004), the adjacent supramarginal gyrus (SMG; Grogan et al., 2012), the inferior frontal gyrus (pars opercularis and triangularis, POp and PTr) (Kaiser et al., 2015; Klein et al., 2014; Mårtensson et al., 2012; Stein et al., 2012) and the STG (Mårtensson et al. 2012), which includes, for instance, the primary auditory cortex and Heschl's gyrus (Ressel et al., 2012). The ROIs are illustrated in Fig. 2A.

Regarding the participant selection, our early and late bilingual participants were highly similar to Kaiser and colleagues' (2015) participants (early vs. late bilinguals) and relatively similar to Klein and colleagues' (2014) participants (simultaneous/early sequential vs. late sequential bilinguals). Based on Kaiser and colleagues' VBM findings, group differences in GM architecture should be

observable bilaterally in fronto-temporal and parietal regions, where late sequential bilinguals are expected to have increased GM volume (possibly due to thicker cortex) compared to early bilinguals. VBM and SBA have, however, been occasionally shown to produce slightly differing results (e.g., Bermudez et al., 2009), bringing into question the extent to which SBA findings can be predicted from the prior VBM observations. Thus, based on Klein and colleagues' (2014) SBA findings in particular, we expect to see between-group differences in cortical thickness in the left IFG, where increased thickness is expected for late vs. early bilinguals.

2 Methods

2.1 Participants

Forty-eight young right-handed adults (21 early Finnish-Swedish bilinguals; 27 late English second language speakers, Finnish as a native language), recruited via university mailing lists and matched for age (early bilinguals: $M = 26.5$ years, $SD = 5.5$, late bilinguals: $M = 27.5$ years, $SD = 4.4$, $t(46) = 0.671$, $p > 0.05$) and gender (early bilinguals: 38 % males, late bilinguals: 55 % males, Fisher's exact test $p > 0.05$), participated in the study. The early bilinguals had learned both Finnish and Swedish before the age of 6 (mean L2 AoA = 2.4, $SD = 1.7$), while the late sequential bilinguals had had no bilingual exposure before school, i.e., before the age of 7 (mean L2 AoA = 9.0, $SD = 1.1$). Both groups reported using their L2 languages actively in their everyday lives; on a weekly basis, the early bilinguals estimated using their L2 on average 40 % of the time and the late bilinguals on average 30 % of the time. Although the groups had different L2 languages (Swedish and English for the early and late bilingual groups, respectively), both Swedish and English belong to the Germanic language branch and thus share many linguistic features (König and Auwera, 2002), ensuring a high comparability between the groups.

In addition to acquiring their L2 languages, both the early and late bilinguals had studied some additional foreign languages within the Finnish basic education system. The first language introduced at school was English at the age of 9 for both groups. For early bilinguals, English is thus their L3 (mean L3 AoA = 9.4 years, SD = 0.7), whereas for the late bilinguals it is L2. Moreover, the late bilinguals started learning the obligatory Swedish as their L3 on average at the age of 11 (mean L3 AoA = 12.4 years, SD = 1.1). Thus, both groups had some knowledge of all three languages (Finnish, Swedish, English), however, all participants estimated L2 to be their strongest¹ and the most used language after Finnish. In total, the early bilingual participants reported having studied, on average, 2.2 (SD = 1.1) post-L2 languages and the late bilinguals 2.7 (SD = 1.1) post-L2 languages. The groups did not differ with regard to the number of additional languages studied (two-tailed $t(46) = 1.467$, $p = 0.149$), however, the difference in their L3 AoA meant that the groups differed with respect to the average number of years spent studying these post-L2 languages ($M = 10.5$ years, $SD = 3.6$ and $M = 7.0$ years, $SD = 3.1$ for the early and late bilinguals, respectively). To account for this difference, L3 AoA was included as a nuisance covariate in all analysis. Thus, in general, the groups were highly similar in terms of their formal language learning backgrounds, but differed considerably with regards to their L2 AoA. As structural differences based on the L2 AoA have been shown to persist into adulthood and remain observable even though other languages are introduced later in life (Hämäläinen et al., 2017; Kaiser et al., 2015), it stands to reason that any emerging

¹ All early bilinguals, as well as the majority of the late bilinguals, estimated their L2 skills to be either “excellent” or “very good” and only a handful ($N=4$) of the late bilingual participants self-rated their L2 skills as “good”. None reported L2 skills below “good”. As the early bilinguals reported continuous use of both L1 and L2 since the childhood, their self-reported language proficiency was considered a reliable estimate of their objective L2 proficiency. To make sure that the late bilinguals’ L2 proficiency indeed matched their high self-reported level, participants’ L2 language skills were objectively evaluated based on their success in a national English exam, taken as part of the Finnish Matriculation Examination at the end of the Finnish upper secondary school at the age of 18. In line with the self-reported language skills, all late bilingual participants had scored within the top 20 % in their examination.

differences between the groups are likely to reflect the difference in the groups' bilingualism status during childhood (hence the nomenclature). Information regarding participants' detailed language learning background, including estimates of proficiency and usage, was collected via self-evaluation questionnaires. This background data is presented in Table 1.

None of the participants reported any neurological or psychiatric disorders nor had any medications that might affect the central nervous system. The participants gave their written informed consent to participate in the experiments. The experiments were performed in accordance with the Declaration of Helsinki. Ethical permission for the experiment was issued by the University of Helsinki Ethical Review Board in the Humanities and Social and Behavioural Sciences.

2.2 MRI acquisition

Participants were scanned with Siemens Skyra 3T MR (Erlangen, Germany) scanner using a standard 32-channel head matrix coil. High-resolution T1 anatomical images were obtained for each participant with a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence with the following parameters: repetition time (TR) 2530 ms, echo time (TE) 3.3 ms, inversion time (TI) 1100 ms, field-of-view 256 mm, voxel size 1 x 1 x 1 mm³, flip angle 7° and number of averages 1. Total scan time was 6 mins 2 sec.

2.3 Surface-based analysis

Surface-based analysis was conducted using Freesurfer software (version 5.3, <http://freesurfer.net>). The T1-weighted images were processed using the standard processing pipeline optimized for 3T scans and the output inspected visually. Briefly, this processing of the images includes brain

extraction (Ségonne et al., 2004), automated Talairach transformation, intensity normalization (Sled et al., 1998), tessellation of the grey matter white matter boundary, automated topology correction (Fischl et al., 2001; Segonne et al., 2007), and surface deformation following intensity gradients determining the grey/white and grey/cerebrospinal fluid borders (Dale et al., 1999; Dale and Sereno, 1993; Fischl and Dale, 2000). Parcellation of the cerebral cortex is based on the gyral and sulcal geometry (Desikan et al., 2006; Fischl et al., 2004) For the whole-brain analyses, the individual cortical models were registered to a spherical atlas based on the individual cortical folding patterns (Fischl et al., 1999) and smoothed using 15 mm FWHM kernel. The statistical analyses were run in QDEC (<https://surfer.nmr.mgh.harvard.edu/fswiki/Qdec>). Age, gender and L3 AoA were used as nuisance covariates. Correction for multiple comparisons was performed using Monte Carlo simulation and two-sided p-values less than 0.05 corrected for two hemispheres were considered statistically significant. For ROI analysis, the average surface output measures were calculated within the cortical regions (Fischl and Dale, 2000) and analysed further with IBM SPSS Statistics 22.0.

2.4 Statistical analysis

A linear mixed model approach for repeated measures analysis of variance was used with pairwise post-hoc tests adjusted for multiple comparisons (Bonferroni). In the model, surface area, volume, thickness and mean curvature values were treated as dependent variables. Similar to the whole-brain level analysis, age, gender and L3 AoA were added as covariates in all models. For area, volume and mean curvature analyses, estimated total intracranial volume was used as an additional covariate to control for confounding effects due to individual variations in brain size. Primary fixed effects were Group (early bilinguals, late bilinguals) and Group by ROI (all 5 ROIs: PTr, POp, SMG, IPC and

STG) by Hemisphere (left, right) interaction. Diagonal covariance matrix was used in all mixed model analyses.

3 Results

The whole-brain level SBA analyses of GM volume, area, thickness and curvature revealed no significant differences between the groups in either direction. With the ROI analysis, significant three-way interactions between Group, ROI, and Hemisphere were revealed for surface area ($F(18) = 300.2$; $p < 0.001$), volume ($F(18) = 267.6$; $p < 0.001$) and mean curvature ($F(18) = 15.77$; $p < 0.001$). Further pairwise comparisons revealed three ROIs with significant between-group differences, namely, the left POp, and the bilateral STG.

Over the left POp, the early bilinguals had significantly extended cortical surface area ($M = 2.0 \times 10^3 \text{ mm}^2$, $SD = 0.32$ and $M = 1.6 \times 10^3 \text{ mm}^2$, $SD = 0.33$, $p < 0.001$ for early and late bilinguals, respectively) and larger cortical volume compared to the late bilinguals ($M = 6.2 \times 10^3 \text{ mm}^3$, $SD = 1.2$ and $M = 4.8 \times 10^3 \text{ mm}^3$, $SD = 1.1$, $p = 0.001$ for early and late bilinguals, respectively). Similarly, significantly extended surface area for early vs. late bilinguals was also observed over the right STG ($M = 4.0 \times 10^3 \text{ mm}^2$, $SD = 0.6$ and $M = 3.6 \times 10^3 \text{ mm}^2$, $SD = 0.6$, $p < 0.001$ for early and late bilinguals, respectively), however, unlike the effect in the left POp, this difference was not accompanied with a corresponding difference in the GM volume ($p = 0.60$). For mean curvature, the effect stemmed from a difference in the left STG, where the late bilinguals had significantly greater surface curvature ($M = 0.128$, $SD = 0.01$) compared to early bilinguals ($M = 0.119$, $SD = 0.009$, $p = 0.009$). No significant group differences were observed in any of the other ROIs. The significant mixed model results are shown in Fig. 2B.

Although the groups were closely matched to minimise any variations in the language background measures other than the L2 AoA, the participants reported somewhat varying estimates for their daily L2 usage. To test whether differences in L2 usage might have contributed to the effects observed, additional correlation analyses between the relevant surface based measures and the overall L2 usage values were run within the ROIs where significant differences were observed. More specifically, L2 usage was correlated with 1) the volume and surface area measures within the left POp, 2) the mean curvature measures within the left STG and 3) the surface area measures within the right STG. None of the correlations were significant ($p > 0.3$ in all correlations).

4 Discussion

The aim of this study was to examine whether L2 AoA might be an important factor in affecting how the brain responds to bilingual exposure. More specifically, we investigated a hypothesis whereby early L2 experience should induce GM modulations that are characterised by predominant changes in cortical surface area, while late L2 acquisition should relate to changes mainly in cortical thickness. To this end, we concurrently compared several markers of cortical surface morphology between early and late sequential bilinguals on both whole-brain level as well as within a priori defined ROIs. Although the groups did not differ from each other at the whole-brain level, the ROI-based approach revealed significant between-group GM morphology differences over three locations, namely, the left IFG and the STG bilaterally.

Our main finding was that, relative to late L2 acquisition, early bilingual experience was associated with significantly extended cortical surface area over the left POp and the right STG. Over the left POp, the difference in surface area was also reflected as a difference in volume between the groups. These findings suggest a relationship between GM morphology and L2 AoA and support the idea

that acquiring an L2 early in childhood, when surface area development is still ongoing (Remer et al., 2017; Tamnes et al., 2017), might lead to extended surface area in locations demanded by the task (Zatorre et al., 2012). Because the left IFG and the left POp in particular have been shown to be involved and even necessary for phonological language processing (Golestani et al., 2011; Gough et al., 2005; Nixon et al., 2004; Price, 2010; Sahin et al., 2009; Zatorre et al., 1996), we propose that the GM modulations over the left POp might be related to accommodating increased phonological processing demands due to learning two languages early in life. This would be consistent with observations that the left auditory-related regions in general tend to exhibit an extended surface area compared to the right homologue areas (Chance et al., 2006; Meyer et al., 2014) – an asymmetry that has been taken to reflect the left hemispheric advantage in extracting temporal information from the incoming auditory stream. In other words, the larger the surface area over the auditory-related regions, the more efficient the auditory processing. Moreover, similar phonology-related structural modulations associated with early bilingualism have also been observed in the underlying WM tracts, in particular along the left direct segment of the arcuate fasciculus (Hämäläinen et al., 2017), which connects the left IFG to the primary auditory processing areas and has been associated with phonological language functions (Catani et al., 2005; Forkel et al., 2014; López-Barroso et al., 2013). This concordance not only provides further evidence in favour of our current interpretation, but it also supports the emerging pattern localising bilingualism-induced structural modulations to the left IFG (Mårtensson et al. 2012; Stein et al. 2012; Klein et al. 2014; Kaiser et al. 2015). However, as functions of the left POp are not limited to phonological processing (see, for instance, Curley et al., 2018; Krams et al., 1998), future research is needed to confirm this postulated relationship between regional surface area and prolonged experience in bilingual phonological processing.

As for the surface area modulations for early vs. late bilinguals over the right STG, at least two interpretations can be provided. First, this finding is partially in line with the remark that early

binuals tend to have bilaterally larger Heschl's gyrus volume compared to monolinguals (Ressel et al., 2012). Since the right Heschl's gyrus has been implied in e.g. spectral sound perception (Schneider et al., 2005), this effect, like the between-group surface area difference over the left POp, could be linked to bilingualism-induced increased phonological processing demands. It is important to remember, however, that the used STG ROI encompassed also the frontal parts of the STG and therefore it is not possible to delineate whether the observed effect is due to areal expansion in Heschl's gyrus in particular or rather reflects expansion of the right STG as a whole. Second, taking the limitations of the used ROI-approach into account, the enlarged cortical surface area over the right STG could be related to the fact that binuals typically exhibit more bilateral functional activations in linguistic tasks in general compared to monolinguals (for review, see e.g. Hull and Vaid, 2007). Along these lines, the current effect could reflect a more bilaterally balanced phonological/auditory (due to STG housing auditory processing areas) language processing configuration for early vs. late binuals. Notably, unlike the surface area difference over the left POp, this STG effect was not associated with a significant concomitant difference in volume between the groups. This demonstrates the benefit of investigating surface-based measures concurrently with volumetric analysis, as some aspects of bilingualism-induced structural GM changes might be overlooked by investigating GM volume differences only.

While the effects discussed above are in line with the first part of the tested hypothesis in linking early L2 exposure to cortical surface area modulations, the validity of the idea of AoA-specific GM modulations as a whole remains ambiguous: contrary to the two previous SBA studies on bilingualism (Klein et al., 2014; Mårtensson et al., 2012), we found no evidence supporting the postulated association between late L2 acquisition and increased cortical thickness. Based on the robust whole-brain level findings of Klein and colleagues' (2014) study in particular, we expected the late L2 speakers to exhibit a thicker cortex in the left IFG relative to the early binuals.

However, although our sample selection was fairly similar to Klein and colleagues' study, compared to their simultaneous bilinguals (L2 AoA = < 3 years, M = 1 year), our early bilingual group had a more varied L2 AoA: participants with L2 AoA up to 5 years (L2 AoA between 0-5 years, M = 2.35, SD = 1.9) were considered early bilinguals. Given that surface area appears to develop rapidly during the first two years of life (Lyall et al., 2015), after which the development speed gradually slows down (Remer et al., 2017), the wider L2 AoA range might have caused our early bilingual group to encompass some participants who already exhibit mainly thickness-driven L2-related modulations (due to L2 AoA \geq 3 according to Klein and colleagues' finding, N = 6). If so, the difference in thickness between the early and late sequential bilinguals is likely to be reduced. Moreover, although structural differences based on the L2 AoA should remain observable in adulthood even when other languages are introduced later in life (Kaiser et al., 2015), another sample-related factor that might have further obscured potential thickness effects concerns the fact that both groups had studied several other languages after acquiring their L2's. As it is not well known to what extent early exposure to two languages affects the neural machinery and learning mechanisms responsible for acquiring additional foreign languages at later ages, it is possible that the additional post-L2 learnt languages might have induced thickness modulations also in the early bilingual group, thus contributing to similar cortical thickness between the groups. Until these issues have been accounted for, our data is insufficient to refute or support the claim that late L2 acquisition leads to thickness-driven GM modulations. More specifically, some critical questions to be address in the future are a) whether simultaneous bilinguals differ from early bilinguals with respect to surface area measures, i.e., what is the exact L2 AoA after which bilingual experience no longer induces surface area modulations? b) whether area-driven vs. thickness-driven modulations might be observable between simultaneous/early bilinguals and late bilinguals in samples who have not learnt additional post-L2 languages, i.e., does learning additional post-L2 languages later in life indeed induce thickness modulations in simultaneous/early bilinguals?

Finally, in addition to the effects over the left POp and the right STG, we observed between-group effects also in the left STG, where the late bilinguals showed significantly greater mean curvature relative to the early bilinguals. However, because cortical curvature has not been previously used to quantify experience-dependent learning (but see Luders et al. 2012 for curvature changes in response to meditation) and its relationship to GM volume is still largely unknown, there is very little scientific background against which this effect could be discussed. Hence, we refrain from speculating this difference further.

To conclude, our study shows that GM morphology is, at least partially, linked to L2 AoA. More specifically, early exposure to two languages appears to lead to extended cortical surface area over the left posterior IFG and the right STG. Based on the anatomical locations of the effects, we propose that these structural GM modulations might reflect the accommodation of increased phonological processing demands related to acquisition and use of two languages from early on. However, caution is warranted in interpreting the current results since the observed effects were localised and not robust enough to survive the whole-brain level analysis. Nevertheless, as the findings show that differences in surface area measures are not always reflected in measures of GM volume, the current study highlights the need to report cortical surface measures together with volume-based data when addressing bilingualism-induced structural modulations. and provides evidence of gained (referring to experience-dependent increase in, e.g., volume or surface area) GM modulations in response to early bilingualism, as opposed to the majority of previous findings linking increased volume and cortical thickness mainly to late L2 acquisition.

Funding

The study was financially supported by Emil Aaltonen Foundation, University of Helsinki 3-year grants, the Academy of Finland (grant #288880 to ML), Kone Foundation (project grant to AL) and Orion Research Foundation, Finnish Medical Foundation and Academy of Finland (#295580) (research grants to JJ).

Notes

Dr. Joutsa discloses travel grants from Abbvie and Orion.

References

- Abutalebi, J., Canini, M., Della Rosa, P.A., Green, D.W., Weekes, B.S., 2015a. The neuroprotective effects of bilingualism upon the inferior parietal lobule: A Structural Neuroimaging Study in Aging Chinese Bilinguals. *J. Neurolinguistics* 33, 3–13. doi:10.1016/j.jneuroling.2014.09.008
- Abutalebi, J., Canini, M., Della Rosa, P.A., Sheung, L.P., Green, D.W., Weekes, B.S., 2014. Bilingualism protects anterior temporal lobe integrity in aging. *Neurobiol. Aging* 35, 2126–2133. doi:10.1016/j.neurobiolaging.2014.03.010
- Abutalebi, J., Guidi, L., Borsa, V., Canini, M., Della Rosa, P.A., Parris, B.A., Weekes, B.S., 2015b. Bilingualism provides a neural reserve for aging populations. *Neuropsychologia* 69, 201–210. doi:10.1016/j.neuropsychologia.2015.01.040
- Ashburner, J., Friston, K.J., 2000. Voxel-Based Morphometry—The Methods. *Neuroimage* 11, 805–821. doi:10.1006/nimg.2000.0582
- Bermudez, P., Lerch, J.P., Evans, A.C., Zatorre, R.J., 2009. Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb. Cortex* 19,

1583–1596. doi:10.1093/cercor/bhn196

Catani, M., Jones, D.K., Ffytche, D.H., 2005. Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.

Chance, S.A., Casanova, M.F., Switala, A.E., Crow, T.J., 2006. Minicolumnar structure in Heschl's gyrus and planum temporale: Asymmetries in relation to sex and callosal fiber number. *Neuroscience* 143, 1041–1050. doi:10.1016/j.neuroscience.2006.08.057

Curley, L.B., Newman, E., Thompson, W.K., Brown, T.T., Hagler, D.J., Akshoomoff, N., Reuter, C., Dale, A.M., Jernigan, T.L., 2018. Cortical morphology of the pars opercularis and its relationship to motor-inhibitory performance in a longitudinal, developing cohort. *Brain Struct. Funct.* 223, 211–220. doi:10.1007/s00429-017-1480-5

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical Surface-Based Analysis. *Neuroimage* 9, 179–194. doi:10.1006/nimg.1998.0395

Dale, A.M., Sereno, M.I., 1993. Improved Localizadon of Cortical Activity by Combining EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach. *J. Cogn. Neurosci.* 5, 162–176. doi:10.1162/jocn.1993.5.2.162

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. doi:10.1016/j.neuroimage.2006.01.021

Fischl, B., 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., Liu, A., Dale, A.M., 2001. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans. Med. Imaging* 20, 70–80. doi:10.1109/42.906426

Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999. High-resolution intersubject averaging and

a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8, 272–84.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E.,

Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M.,

2004. Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22.

Forkel, S.J., de Schotten, M.T., DellAcqua, F., Kalra, L., Murphy, D.G.M., Williams, S.C.R., Catani,

M., 2014. Anatomical predictors of aphasia recovery: a tractography study of bilateral

perisylvian language networks. *Brain* 137, 2027–2039.

García-Pentón, L., Fernández García, Y., Costello, B., Duñabeitia, J.A., Carreiras, M., 2015. The

neuroanatomy of bilingualism: how to turn a hazy view into the full picture. *Lang. Cogn.*

Neurosci. 3798, 1–25. doi:10.1080/23273798.2015.1068944

Gold, B.T., Johnson, N.F., Powell, D.K., 2013. Lifelong bilingualism contributes to cognitive reserve

against white matter integrity declines in aging. *Neuropsychologia* 51, 2841–2846.

Golestani, N., Price, C.J., Scott, S.K., 2011. Born with an Ear for Dialects? Structural Plasticity in

the Expert Phonetician Brain. *J. Neurosci.* 31, 4213–4220. doi:10.1523/JNEUROSCI.3891-

10.2011

Gough, P.M., Nobre, A.C., Devlin, J.T., 2005. Dissociating Linguistic Processes in the Left Inferior

Frontal Cortex with Transcranial Magnetic Stimulation. *J. Neurosci.* 25, 8010–8016.

doi:10.1523/JNEUROSCI.2307-05.2005

Grogan, A., 'O. Parker Jones, T., Ali, N., Crinion, J., Orabona, S., Mechias, M.L., Ramsden, S.,

Green, D.W., Price, C.J., 2012. Structural correlates for lexical efficiency and number of

languages in non-native speakers of English. *Neuropsychologia* 50, 1347–1352.

doi:10.1016/j.neuropsychologia.2012.02.019

Hull, R., Vaid, J., 2007. Bilingual language lateralization: A meta-analytic tale of two hemispheres.

Neuropsychologia 45, 1987–2008.

Hämäläinen, S., Sairanen, V., Leminen, A., Lehtonen, M., 2017. Bilingualism modulates the white

matter structure of language-related pathways. *Neuroimage* 152, 249–257.

doi:10.1016/j.neuroimage.2017.02.081

Im, K., Lee, J.-M., Won Seo, S., Hyung Kim, S., Kim, S.I., Na, D.L., 2008. Sulcal morphology changes and their relationship with cortical thickness and gyral white matter volume in mild cognitive impairment and Alzheimer's disease. *Neuroimage* 43, 103–113.

doi:10.1016/j.neuroimage.2008.07.016

Kaiser, A., Eppenberger, L.S., Smieskova, R., Borgwardt, S., Kuenzli, E., Radue, E.W., Nitsch, C., Bendfeldt, K., 2015. Age of second language acquisition in multilinguals has an impact on gray matter volume in language-associated brain areas. *Front. Psychol.* 6, 638.

doi:10.3389/fpsyg.2015.00638

Klein, D., Mok, K., Chen, J.K., Watkins, K.E., 2014. Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain Lang.* 131, 20–24. doi:10.1016/j.bandl.2013.05.014

Krams, M., Rushworth, M.F.S., Deiber, M.-P., Frackowiak, R.S.J., Passingham, R.E., 1998. The preparation, execution and suppression of copied movements in the human brain. *Exp. Brain Res.* 120, 386–398. doi:10.1007/s002210050412

König, E., Auwera, J. van der., 2002. *The Germanic languages*. Routledge.

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

López-Barroso, D., Catani, M., Ripollés, P., Dell'Acqua, F., Rodríguez-Fornells, A., de Diego-Balaguer, R., 2013. Word learning is mediated by the left arcuate fasciculus. *Proc. Natl. Acad. Sci.* 110, 13168–13173.

Luders, E., Kurth, F., Mayer, E.A., Toga, A.W., Narr, K.L., Gaser, C., 2012. The Unique Brain Anatomy of Meditation Practitioners: Alterations in Cortical Gyrfication. *Front. Hum. Neurosci.* 6, 34. doi:10.3389/fnhum.2012.00034

- Lyall, A.E., Shi, F., Geng, X., Woolson, S., Li, G., Wang, L., Hamer, R.M., Shen, D., Gilmore, J.H., 2015. Dynamic Development of Regional Cortical Thickness and Surface Area in Early Childhood. *Cereb. Cortex* 25, 2204–2212. doi:10.1093/cercor/bhu027
- Mechelli, A., Crinion, J.T., O’Doherty, U.N.J., Ashburner, J., R S Frackowiak, S.R., Price, C.J., 2004. Neurolinguistics: structural plasticity in the bilingual brain. *Nature* 431, 757.
- Meyer, M., Liem, F., Hirsiger, S., J?ncke, L., H?nggi, J., 2014. Cortical surface area and cortical thickness demonstrate differential structural asymmetry in auditory-related areas of the human cortex. *Cereb. Cortex* 24, 2541–2552. doi:10.1093/cercor/bht094
- Mårtensson, J., Eriksson, J., Bodammer, N.C., Lindgren, M., Johansson, M., Nyberg, L., Lövdén, M., 2012. Growth of language-related brain areas after foreign language learning. *Neuroimage* 63, 240–244. doi:10.1016/j.neuroimage.2012.06.043
- Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., Passingham, R., 2004. The Inferior Frontal Gyrus and Phonological Processing: An Investigation using rTMS. *J. Cogn. Neurosci.* 16, 289–300. doi:10.1162/089892904322984571
- Olsen, R.K., Pangelinan, M.M., Bogulski, C., Chakravarty, M.M., Luk, G., Grady, C.L., Bialystok, E., 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
- Pakkenberg, B., Gundersen, H.J., 1997. Neocortical neuron number in humans: effect of sex and age. *J. Comp. Neurol.* 384, 312–20.
- Panizzon, M.S., Fennema-Notestine, C., Eyler, L.T., Jernigan, T.L., Prom-Wormley, E., Neale, M., Jacobson, K., Lyons, M.J., Grant, M.D., Franz, C.E., Xian, H., Tsuang, M., Fischl, B., Seidman, L., Dale, A., Kremen, W.S., 2009. Distinct genetic influences on cortical surface area and cortical thickness. *Cereb. Cortex* 19, 2728–35. doi:10.1093/cercor/bhp026
- Pliatsikas, C., Johnstone, T., Marinis, T., 2014. Grey matter volume in the cerebellum is related to the processing of grammatical rules in a second language: A structural voxel-based

morphometry study. *Cerebellum* 13, 55–63. doi:10.1007/s12311-013-0515-6

Price, C.J., 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. N. Y. Acad. Sci.* 1191, 62–88. doi:10.1111/j.1749-6632.2010.05444.x

Rakic, P., 2009. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* 10, 724–735. doi:10.1038/nrn2719

Raznahan, A., Shaw, P., Lalonde, F., Stockman, M., Wallace, G.L., Greenstein, D., Clasen, L., Gogtay, N., Giedd, J.N., 2011. How does your cortex grow? *J. Neurosci.* 31, 7174–7. doi:10.1523/JNEUROSCI.0054-11.2011

Remer, J., Croteau-Chonka, E., Dean, D.C., D'Arpino, S., Dirks, H., Whiley, D., Deoni, S.C.L., 2017. Quantifying cortical development in typically developing toddlers and young children, 1–6 years of age. *Neuroimage* 153, 246–261. doi:10.1016/j.neuroimage.2017.04.010

Ressel, V., Pallier, C., Ventura-Campos, N., Díaz, B., Roessler, A., Ávila, C., Sebastián-Gallés, N., 2012. An effect of bilingualism on the auditory cortex. *J. Neurosci.* 32, 16597–601. doi:10.1523/JNEUROSCI.1996-12.2012

Sahin, N.T., Pinker, S., Cash, S.S., Schomer, D., Halgren, E., 2009. Sequential Processing of Lexical, Grammatical, and Phonological Information Within Broca's Area. *Science* (80-.). 326.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H.J., Dosch, H.G., Bleeck, S., Stippich, C., Rupp, A., 2005. Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nat. Neurosci.* 8, 1241–1247. doi:10.1038/nn1530

Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22, 1060–1075. doi:10.1016/j.neuroimage.2004.03.032

Segonne, F., Pacheco, J., Fischl, B., 2007. Geometrically Accurate Topology-Correction of Cortical Surfaces Using Nonseparating Loops. *IEEE Trans. Med. Imaging* 26, 518–529. doi:10.1109/TMI.2006.887364

- Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17, 87–97.
doi:10.1109/42.668698
- Stein, M., Federspiel, A., Koenig, T., Wirth, M., Strik, W., Wiest, R., Brandeis, D., Dierks, T., 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
- Stein, M., Winkler, C., Kaiser, A.C., Dierks, T., 2014. Structural brain changes related to bilingualism: Does immersion make a difference? *Front. Psychol.* 5, 1–7.
doi:10.3389/fpsyg.2014.01116
- Tamnes, C.K., Herting, M.M., Goddings, A.-L., Meuwese, R., Blakemore, S.-J., Dahl, R.E., Güroğlu, B., Raznahan, A., Sowell, E.R., Crone, E.A., Mills, K.L., 2017. Development of the Cerebral Cortex across Adolescence: A Multisample Study of Inter-Related Longitudinal Changes in Cortical Volume, Surface Area, and Thickness. *J. Neurosci.* 37, 3402–3412.
doi:10.1523/JNEUROSCI.3302-16.2017
- Wallace, G.L., Eisenberg, I.W., Robustelli, B., Dankner, N., Kenworthy, L., Giedd, J.N., Martin, A., 2015. Longitudinal Cortical Development During Adolescence and Young Adulthood in Autism Spectrum Disorder: Increased Cortical Thinning but Comparable Surface Area Changes. *J. Am. Acad. Child Adolesc. Psychiatry* 54, 464–469. doi:10.1016/j.jaac.2015.03.007
- Wierenga, L.M., Langen, M., Oranje, B., Durston, S., 2014. Unique developmental trajectories of cortical thickness and surface area. *Neuroimage* 87, 120–126.
doi:10.1016/j.neuroimage.2013.11.010
- Zatorre, R.J., Fields, R.D., Johansen-Berg, H., 2012. Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nat. Neurosci.* 15, 528–536. doi:10.1038/nn.3045
- Zatorre, R.J., Meyer, E., Gjedde, A., Evans, A.C., 1996. PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cereb Cortex* 6, 21–30. doi:10.1093/cercor/6.1.21

Zielinski, B.A., Prigge, M.B.D., Nielsen, J.A., Froehlich, A.L., Abildskov, T.J., Anderson, J.S., Fletcher, P.T., Zygmont, K.M., Travers, B.G., Lange, N., Alexander, A.L., Bigler, E.D., Lainhart, J.E., 2014. Longitudinal changes in cortical thickness in autism and typical development. *Brain* 137, 1799–1812. doi:10.1093/brain/awu083

Tables and Figure Captions

Table 1.		
Participants' language background.		
Age of acquisition	Early bilinguals, M (SD)	Late bilinguals, M (SD)
Finnish	from birth	from birth
Swedish	2.35 (1.9)	12.4 (1.1)
English	9.35 (0.7)	8.96 (1.1)
Languages used during childhood	% Early bilinguals	% Late bilinguals
Both languages at home	70.6	-
FIN at home, SWE outside home	17.6	-
SWE at home, FIN outside home	11.8	-
Only FIN	-	100
Post-L2 language history	Early bilinguals, M (SD)	Late bilinguals, M (SD)
No. of post-L2 lang. studied	2.2 (1.1)	2.7 (1.1)
Years of studying post-L2 lang.	10.5 (3.6)	7.0 (3.1)
Language usage	Early bilinguals, mean %	Late bilinguals, mean %
	L1 / L2 / Other	L1 / L2 / Other
Overall	40.0 / 37.1 / 22.4	68.3 / 28.3 / 3.3
Speaking	43.5 / 44.1 / 12.4	85.8 / 13.8 / 0.8
Reading	30.6 / 37.6 / 31.8	53.3 / 43.3 / 3.3
Listening	41.8 / 33.5 / 24.7	62.9 / 34.6 / 2.5
Writing	30.6 / 45.9 / 25.0	60.8 / 36.7 / 2.9
Self-estimated language skills ^a	Early bilinguals, Median	Late bilinguals, Median
	L1 / L2 / Other	L1 / L2 / Other
Overall	7 / 7 / 6	7 / 7 / 5
Speaking	7 / 7 / 6	7 / 6 / 5
Reading	7 / 7 / 6	7 / 6 / 5
Listening	7 / 7 / 6	7 / 6 / 5
Writing	7 / 7 / 6	7 / 6 / 5
Objective language skills		Late bilinguals, Median
Grade in the English part of the national Finnish Matriculation Examination		E ^b

^a A seven-point Likert-type scale, where 7=Excellent, 6=Very Good, 5=Good, 4=Average, 3=Acceptable, 2=Poor, 1=Very Poor

^b E = eximia cum laude approbatur, the second best grade, top 20% of students

Fig. 1. A hypothetical situation where GM volume (blue and red cubes) appears identical between the groups despite significant differences in cortical surface area (dashed line) and thickness (solid line).

Fig. 2. A) Illustration of the included ROIs: POp in light brown, PTr in orange, SMG in green, IPC in purple and STG in cyan portrayed on a standard MNI152 brain. The colouring of the ROIs is adapted from Desikan and colleagues' (2006) original figure. **B)** Cortical area, volume and mean curvature values from ROIs where significant differences between groups were observed with the linear mixed model analysis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. POp = pars opercularis, PTr = pars triangularis, SMG = supramarginal gyrus, IPC = inferior parietal cortex, STG = superior temporal gyrus. Note that area is in units of $\times 10^3 \text{ mm}^2$ and volume is in units of $\times 10^3 \text{ mm}^3$, mean curvature is unitless.