Bilingualism: Language and Cognition

cambridge.org/bil

White matter microstructure predicts foreign language learning in army interpreters

Johan Mårtensson[1], Johan Eriksson[2], Nils Christian Bodammer[3], Magnus Lindgren[4], Mikael Johansson[5], Lars Nyberg[6] and Martin Lövdén[7]

[1]Faculty of Medicine, Department of Clinical Sciences Lund, Logopedics, Phoniatrics and Audiology, Lund University, 22100 Lund, Sweden; [2]Umeå Center for Functional Brain Imaging, Umeå University, 90187 Umeå, Sweden; [3]Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany; [4]Department of Psychology, Lund University, 22100 Lund, Sweden; [5]Department of Psychology, Lund University, 22100 Lund, Sweden; [6]Umeå Center for Functional Brain Imaging, Umeå University, 90187 Umeå, Sweden and [7]Aging Research Center, Karolinska Institutet, Gävlegatan 16, 11330 Stockholm, Sweden

Research Article

Cite this article: Mårtensson J, Eriksson J, Bodammer NC, Lindgren M, Johansson M, Nyberg L, Lövdén M (2020). White matter microstructure predicts foreign language learning in army interpreters. *Bilingualism: Language and Cognition* **23**, 763–771. https://doi.org/10.1017/S1366728920000152

Received: 2 July 2018
Revised: 28 January 2020
Accepted: 29 January 2020
First published online: 3 March 2020

Keywords:
diffusion tensor imaging; language acquisition; interpreting

Address for correspondence:
Johan Mårtensson,
E-mail: johan.martensson@med.lu.se

Abstract

Adult foreign language acquisition is challenging, and the degree of success varies among individuals. Anatomical differences in brain structure prior to training can partly explain why some learn more than others. We followed a sample of conscript interpreters undergoing intense language training to study learning-related changes in white-matter microstructure (FA, MD, RD and AD) and associations between differences in brain structure prior to training with acquired language proficiency. No evidence for changes in white matter microstructure relative to a control group was found. Starting values of RD, AD and MD were positively related to final test scores of language proficiency, corroborating earlier findings in the field and highlighting the need for further study of how initial brain structure influences and interacts with learning outcomes.

Introduction

Adult second language acquisition is demanding (Sakai, 2005) and individuals vary in terms of their learning efficacy (Qi, Han, Garel, Chen & Gabrieli, 2015). Several studies have highlighted the importance of structural and functional brain connectivity for individual differences in learning capacity (Xiang, Dediu, Roberts, Oort, Norris & Hagoort, 2012; Sheppard, Wang & Wong, 2012; Hosoda, Tanaka, Nariai, Honda & Hanakawa, 2013; Lopez-Barroso, de Diego-Balaguer, Cunillera, Camara, Münte & Rodriguez-Fornells, 2011; Qi et al., 2015 and Ripollés, Biel, Peñaloza, Kaufmann, Marco-Pallarés, Noesselt & Rodríguez-Fornells, 2017). There is also increasing evidence that predispositions in brain structure (and function) play a part for music and speech learning (Zatorre, 2013) as well as language acquisition (Qi et al., 2015) and more generally as a determinant for acquiring expertise (Ullén, Hambrick & Mosing, 2016).

Functional and structural neuroimaging can help explain why some learn better than others. Yang, Gates, Molenaar and Li (2015) used functional Magnetic Resonance Imaging (fMRI) to study native English speakers as they were taught a new tonal vocabulary over the course of six weeks. Their results show that successful and less successful learners differed in brain activation prior to training, with greater neural activity during tone discrimination for successful learners, and more activity during pitch discrimination for less successful learners. Over time successful learners also showed a more coherent and integrated functional network. Shepard, Wang and Wong (2012) used graph theory to look at functional network characteristics (using fMRI) in relation to learning outcomes following an auditory pitch discrimination task. They found that successful learners showed reduced local efficiency but increased global efficiency in a core network of auditory language areas.

Other studies have combined measures of functional (using fMRI) and structural (using Diffusion Tensor Imaging: DTI) brain connectivity. López-Barrosso and colleagues (2011) showed that efficient word learners rely on fast and efficient communication between temporal and frontal areas in the left hemisphere whilst more recent findings (Ripollés et al., 2017) have linked brain connectivity in temporal pathways to correct identification of words and meanings. Qi et al. (2015) compared DTI measurements of known language tracts to the outcome of four weeks of Mandarin training and saw that language learning was correlated with DTI measures in the right hemisphere, but not in the left. These findings, taken together with work by Xiang et al. (2012), show that individual differences in brain connectivity can help explain why some people learn a language more successfully than others.

Structural brain change (in the form of alterations in grey and white matter microstructure) can also occur as an effect of learning a new language (see Li, Legault and Litcofsky (2014) for

© The Author(s) 2020. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CAMBRIDGE UNIVERSITY PRESS

review). We identified local increases in grey matter volume as an effect of learning a new language (Mårtensson, Eriksson, Bodammer, Lindgren, Johansson, Nyberg & Lövdén, 2012). Intense language learning in a group of conscript interpreters at the Swedish Armed Forces Language School led to regional changes in cortical thickness as well as changes in right hippocampal volume. Importantly, these changes were related to learning outcomes, with larger increases in right hippocampal volume and the superior temporal gyrus in interpreters who became more proficient in their assigned language. White matter microstructure has also been known to change in response to language training. Schlegel, Rudelson and Tse (2012) collected monthly DTI scans of English speakers who underwent a 9-month intensive course in Modern Standard Chinese and saw that white matter networks reorganized progressively during that time period. Hosoda and colleagues (2013) investigated a larger cohort of participants (n 137) using both grey matter and white matter measures. They found that grey matter structure was initially related to vocabulary competence but did not predict the outcome of a 16-week 2nd language training program. They did, however, see changes in both white matter microstructure and regional grey matter volume, primarily in the right hemisphere, as an effect of training. Language studies can also lead to brain changes on the functional level (Paulesu, Vallar, Berlingeri, Signorini, Vitali, Burani, Perani & Fazio, 2009; e.g., Shtyrov, 2011; Shtyrov, Nikulin & Pulvermüller, 2010).

The highlighted studies point towards the importance of the language network as a determinant for later language performance. And even more importantly, they illustrate the ability for said network to change in response to demands. The present study follows a select group of multilinguals as their language networks are pushed at a high pace, allowing us to observe how an experienced brain responds to training. Studying this group can lend valuable insight into how learning affects other advanced groups that learn multiple languages over the course of their careers, such as interpreters or diplomats. Our earlier findings (Mårtensson et al., 2012) were limited to grey matter structure. In light of the functional and structural findings presented above it is relevant to investigate whether a) white matter and grey matter microstructure will act as a predictor of learning outcomes and b) whether white matter microstructure will change in response to language training in a group of multilingual individuals who were pre-selected for their language ability prior to training.

Grey matter and white matter measurements from the same population as in Mårtensson et al. (2012) were used for this study. Individuals with strong language ability and knowledge in at least 2 foreign languages enrolled at the Swedish Armed Forces Language School and studied at a pace of 350-500 new words per week. Increases in right hippocampal volume and changes in cortical thickness of the inferior frontal gyrus (IFG), superior temporal gyrus (STG) and the middle frontal gyrus (MFG) were observed following 3 months of language learning. These areas hold key roles in sensorimotor aspects of language (Demonet, 2005; Hickok & Poeppel, 2007; Price, 2010). More specificically, the IFG is belived to be involved in the articulatory network (predominantly left hemisphere) whilst the STG is concerned with spectrotemporal analysis (bilaterally). Both belong to the same frontal network of areas regulating speech processing (Hickok & Poeppel, 2007). The hippocampus is also believed to be involved in rapid learning of new words (Davis & Gaskell, 2009).

There are known connections (see Friederici (2009), for review) between the IFG and the STG, such as the Arcuate Fasciculus and the Superior Longitudinal Fasciculus (SLF). The arcuate fasciculus forms parts of the SLF and has traditionally been claimed to connect Broca's and Wernicke's areas (Catani & Thiebaut de Schotten, 2008), a statement that has been challenged as advancements in neuroimaging have taken place (Bernal & Altman, 2010). Another pathway that is believed to be related to language is the uncinate fasciculus. It connects the anterior temporal lobe with the medial and lateral orbitofrontal cortex (Catani & Thiebaut de Schotten, 2008).

Grey matter consists primarily of neuronal cell bodies, dendrites, and axon collaterals along with glia (mainly astrocytes), synapses, and capillaries whilst white matter is dominated by myelinated axons (Kassem, Lagopoulos, Stait-Gardner, Price, Chohan, Arnold, Hatton & Bennett, 2012; Zatorre, Fields & Johansen-Berg, 2012). White matter microstructure is often measured using Diffusion-tensor imaging (DTI,). DTI is sensitive to hindrance of water diffusion that results from tissue boundaries and is often quantified using mean diffusivity (MD), fractional anisotropy (FA), radial diffusivity (RD) and axial diffusivity (AD). MD quantifies free diffusion of water within a voxel (Beaulieu, 2002) whilst FA measures the directionality of diffusion. FA and MD are calculated from AD and RD, which measure diffusion parallel to and perpendicular to axonal fibers, respectively (Sen & Basser, 2005).

Considering earlier studies from the field, and the findings by Mårtensson et al. (2012) specifically, we hypothesize that grey and white matter will be predictive of later learning outcomes as measured by language proficiency and that white matter microstructure will change in response to training. Left and right hippocampal volumes are expected to be predictive of later language proficiency, as are areas in the language network (Friederici, 2011; Friederici & Gierhan, 2013). Brain structure will be measured globally using Tract-Based Spatial Statistics (TBSS: white matter) and using FreeSurfer's cortical stream (grey matter). Changes are expected to be larger in, but not restricted to, the left hemisphere and with projections towards the right temporal lobe in light of the findings of Mårtensson et al. (2012).

Methods

Participants

Fourteen (6 female) right-handed and MRI-eligible volunteers were recruited from the Swedish Armed Forces Intelligence and Security Centre. Conscripts at the center were selected from all Swedish 18-year-old men and women who had willfully decided to undergo military training. Screening before entry was based on school achievements, study skills, emotional stability, and intelligence. Students are required to have top grades in at least 2 foreign languages (making all students at least tri-lingual) and are given a week to learn 350 non-words of Finnish origin. The purpose of this vocabulary test is to select recruits that can manage the very high pace of the academy. Out of roughly a hundred select applicants the academy chooses between 20-30 students in an average year (see Dahlquist, 2004 for reference). Eight of the interpreters studied Dari, four studied Arabic, and two studied Russian. No interpreter had prior knowledge in his or her assigned language.

Controls (n = 17, 10 female) were students of cognitive science or medicine at Umeå University. The control group was recruited

https://doi.org/10.1017/S1366728920000152 Published online by Cambridge University Press

to be comparable to interpreters in age, intelligence, years of education, and emotional stability. Controls were measured before and their semester, which matched closely in time to the measurements of interpreters. The time between measurements was 3 months.

The groups were equivalent in age, years of education, intelligence and emotional stability. See Table 1.

Behavioral measures

Raven's advanced progressive matrices

The 18 odd-numbered items from set II of this test (*Raven, 2000*) were administered to both groups at pretest. Participants had 10 minutes to complete the task, and the dependent variable was the number of correctly selected patterns.

Anxiety ratings

A Swedish translation of the STAI Y-2 (Spielberger, Gorsuch & Lushene, 1970) was modified and presented to participants who rated anxiety (Filaire, Sagnol, Ferrand, Maso & Lac, 2001) they had experienced in the past month. The questionnaire was administered at pretest.

Proficiency

Our proficiency measure consisted of grades from the mid-year exam at the interpreter academy. This test was performed a few weeks after posttest. This exam is especially important to the interpreters because those who fail are forced to leave the academy (none of our participants had to leave, indicating that they studied hard). The exam itself consists of one written and one oral test. The written language test includes translating full sentences and texts and the oral test includes non-simultaneous interpreting. Both tests have been developed to measure the ability of actual language use in the demanding circumstances a military interpreter might find herself in, and as such they measure a broad spectrum of language abilities. For our proficiency measure we used the means of the oral and written exam, with a scale between 1 (least proficient) and 10 (most proficient). During the stay at the academy the interpreters underwent similar tests, with oral and written exams interleaved (one per week). The teachers had no insight into the findings from the study when grading the students. The tests were developed at the academy (which has been active since 1957) by language teachers, most of whom work part time as lecturers at Swedish universities.

Struggle

To investigate whether training-induced white matter changes were connected to increased knowledge or/and a large amount of effort, we asked the head teacher at the academy to subjectively rate the amount of effort needed to stay at the academy at post-test. The question (translated from Swedish) was: "Judge how large effort was needed for each participant to achieve the goals of the interpreter academy and to be allowed to stay in the program" and was rated on a Likert scale of 1 (little effort) to 9 (large effort). No participant scored below 6 on this scale which led to a restriction in range.

Known languages

All participants knew at least two languages (Swedish and English) with additional languages known by both interpreters and controls. In effect, all participants were at least trilingual, which is of relevance since learning and using a second language can affect white matter microstructure (Pliatsikas, Moschopoulou & Saddy, 2015). Statistically the groups did not differ in terms of number of languages known (see Table 1.). However, it is difficult to rule out the possibility that there actually might have been differences between the groups, since the participants were simply asked to note down any additional languages they know (aside from Swedish and English) and not the language ability in each additional language.

MR Acquisition

Images were acquired at Umeå center for Functional Brain Imaging (UFBI) on a GE Discovery MR 750, 3 Tesla scanner with a 32-channel phased-array head coil. For Diffusion Weighted Images at pre- and post-test a Spin Echo refocused EPI sequence was used. Images had a slice thickness of 2 mm. Sequence parameters were: TR = 8000 ms, 64 slices with no gap, acquisition matrix 128×128 interpolated to 256×256 matrix with a FOV of 250 mm, TE= 84.4 ms, 4 repetitions of 24 independent directions, b= 1000 s/mm2 and 4 b = 0 images, Dual Spin Echo switched on and ASSET acceleration factor 2.

For T1-weighted imaging a 3D fast spoiled gradient-echo (fSPGR) sequence was used at pretest and post-test (TE = 3.2 ms, TR = 8.1 ms, TI = 450 ms, flip angle = 12°, FOV = 172×250×250 mm3, matrix = 172×256×256, bandwidth per pixel = 122 Hz, no parallel imaging, no surface coil intensity correction, 3D correction for gradient non-linearities).

DTI Preprocessing

Diffusion-weighted images were analyzed using the FSL software package (http://www.fmrib.ox.ac.uk/fsl). The four subject-specific images were averaged to improve signal-to-noise ratio and the resulting output image was corrected for possible head movement. This was done using FLIRT from FSL (Jenkinson & Smith, 2001; Jenkinson, Bannister, Brady & Smith, 2002) with the mean of the B = 0 images from each run being used as the reference using 6 degrees of freedom followed by interpolation using nearest neighbor. Images from all participants were inspected for motion and no participant was excluded. A mean image was calculated from the B = 0 image of each run and used as a brain mask. The resulting data was then processed via dtifit and voxelwise statistical analysis was carried out using TBSS v1.2 (Smith, Jenkinson, Johansen-Berg, Rueckert, Nichols, Mackay, Watkins, Ciccarelli, Cader, Matthews & Behrens, 2006; Smith, Jenkinson, Woolrich, Beckmann, Behrens, Johansen-Berg, Bannister, De Luca, Drobnjak, Flitney & Niazy, 2004): all FA images were aligned into 1×1x1mm standard space with the FMRIB58_FA image as target; the most typical subject (the participant that corresponds the most to the rest of the sample) was then used for group-wide alignment into standard space; the resulting image was fed into a tract skeleton generation program and a threshold set at 0.2 to exclude gray matter voxels or cerebrospinal fluid. The MD, RD and AD images made use of the non-linear transformation and projection vectors extracted during preprocessing of the FA images but were otherwise treated in the same way as above.

Analysis of DTI data

Difference images

For each participant, a difference image was calculated from the skeletonized images (by subtracting the pretest image from the

https://doi.org/10.1017/S1366728920000152 Published online by Cambridge University Press

Table 1. Means and Standard Deviations for behavioral measures

	Interpreters		Controls		
Measurement	Mean	SD	Mean	SD	*t*-test
Chronological age	19.9	*1.8*	20.6	*1.0*	$t(28)=-1.34$, $p=.19$
Years of education	12.9	*1.6*	13.3	*0.9*	$t(28)=-0.93$, $p=.36$
Ravens advanced progressive matrices	10.8	*2.9*	11.0	*4.4*	$t(28)=-0.16$ $p=.88$
STAI Y-2	38	*5.9*	34.8	*6.6*	$t(28)=1.38$, $p=.18$
Known languages[1]	2.13	0.31	1.53	*0.286*	*$t(28)=1.44$, $p=.16$*
Proficiency[2]	7	1.89			
Struggle[2]	7.29	1.33			

[1]Additional languages known on top of Swedish and English.
[2]Measures of Proficiency and Struggle were only collected for the Interpreter population.

posttest image). This was done for FA, MD, RD and AD values respectively. Interpreters were then compared against controls on these difference images by means of voxelwise permutation-based inference (5000 permutations). Significant differences in this test thus reflect differences between the groups in the amount of changes between pretest and posttest (i.e., a group by time interaction). Threshold-Free Cluster Enhancement (TFCE) was used and only TFCE p-value images, fully corrected for multiple comparisons across space and at $p < 0.05$ were considered.

Pretest differences

Due to the select nature of the interpreters as compared to the control population we compared FA, MD, RD and AD between groups at pretest using the criteria above.

Behavioral correlations to change

To measure whether difference images correlated with either Struggle or Proficiency in the interpreters a third analysis was carried out, using the difference images but with Struggle and Proficiency added as covariates (in two separate runs).

Predictive value of white matter microstructure

A fourth analysis was performed to evaluate whether FA, MD, RD or AD values at pretest were related to language proficiency at posttest. Within the group of interpreters, demeaned proficiency values were added as a covariate to the same type of analysis and then tested against a null distribution generated from 5000 iterations. The mean values from the resulting skeletonized MD, RD and AD values were then exported (using fslstats with the -M option) into SPSS version 26 (IBM Corp, 2019) for outlier analysis (using the SPSS boxplot feature, which highlights outliers based on Interquartile range of 1.5), and comparison to prior grey-matter findings from Mårtensson et al. (2012). A second run was performed to account for measures of intelligence.

Pre-test analysis of FreeSurfer data from Mårtensson et al. (2012).

To provide completion in light of earlier findings that found grey matter to be predictive in a similar population (Hosoda et al., 2013) the published cortical and hippocampal data from the presented cohort was revisited to investigate whether cortical thickness or hippocampal volume was predictive of later language proficiency. All scans were manually inspected for artifacts and no scan was rejected. The volumes were then analyzed using the FreeSurfer imaging analysis suite (http://surfer.nmr.mgh.harvard.edu/; version 4.5). FreeSurfer makes use of intensity and continuity information from MR volumes to reconstruct and measure cortical thickness and volumetric segmentation in a separate analysis stream to measure subcortical structures. The technical details of these procedures have been described in earlier publications (Dale, Fischl & Sereno, 1999; Dale & Sereno, 1993; Desikan, Segonne, Fischl, Quinn, Dickerson, Blacker, Buckner, Dale, Maguire, Hyman, Albert & Killiany, 2006; Fischl & Dale, 2000; Fischl, Liu & Dale, 2001; Fischl, Salat, Busa, Albert, Dieterich, Haselgrove, van der Kouwe, Killiany, Kennedy, Klaveness, Montillo, Makris, Rosen & Dale, 2002; Fischl, Salat, van der Kouwe, Makris, Segonne, Quinn & Dale, 2004; Fischl, van der Kouwe, Destrieux, Halgren, Segonne, Salat, Busa, Seidman, Goldstein, Kennedy, Caviness, Makris, Rosen & Dale, 2004; Fischl, Sereno & Dale, 1999; Fischl, Sereno, Tootell & Dale, 1999; Han, Jovicich, Salat, van der Kouwe, Quinn, Czanner, Busa, Pacheco, Albert, Killiany, Maguire, Rosas, Makris, Dale, Dickerson & Fischl, 2006; Jovicich, Czanner, Greve, Haley, van der Kouwe, Gollub, Kennedy, Schmitt, Brown, Macfall, Fischl & Dale, 2006; Segonne, Dale, Busa, Glessner, Salat, Hahn & Fischl, 2004).

Vertex-wise general linear model analysis was performed with Proficiency as the dependent variable and cortical thickness as the independent variable. For subcortical grey matter structure, a linear regression was performed using Proficiency as the dependent variable and right and left hippocampal volume as independent variables. See Mårtensson et al. (2012) for further details, and the results below.

Results

Pretest differences in white matter microstructure (RD) between interpreters and controls but no change over time

No statistically significant pretest differences were found between controls and interpreters for FA, MD or AD. However, interpreters had relatively lower RD values bilaterally in several areas involved in the language network (Friederici, 2011; Friederici & Gierhan, 2013): forceps minor, inferior fronto-occipital fasciculus, superior longitudinal fasciculus, uncinate fasciculus, anterior thalamic radiation, inferior longitudinal fasciculus with a larger area visible in the left hemisphere, the corticospinal tract as well as the cingulum (see Figure 1A). Controls did not have lower RD values compared to interpreters in any area.

https://doi.org/10.1017/S1366728920000152 Published online by Cambridge University Press

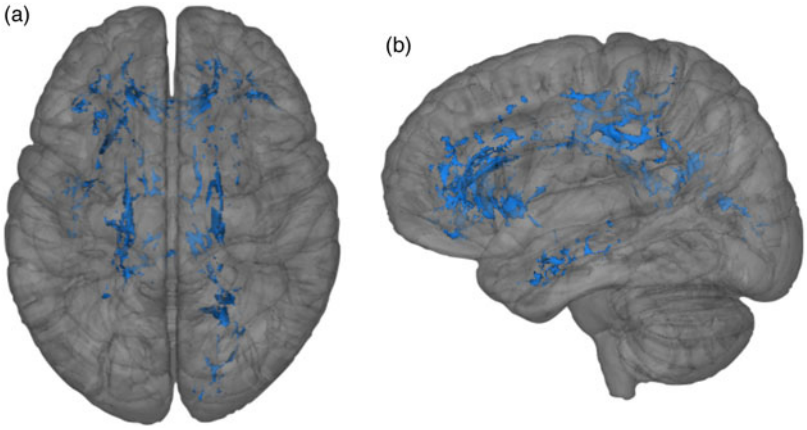


Fig. 1. Interpreters exhibited lower radial diffusivity values at baseline as compared to controls. Results as seen from the top (A) and from the left (B).

Calculated difference images (post – pre) showed no selective changes over time for FA, MD, RD or AD for interpreters relative to controls. Hence, no evidence for changes in white matter microstructure over time were found.

An additional analysis compared RD at posttest between both groups to see whether the initial difference between the populations was still present. There was no difference between the groups for RD at posttest (at $p < 0.05$) but a tendency ($p > 0.1$) for interpreters to have lower RD in the same networks at posttest as at pretest, again indicating that no real change between the groups occurred as an effect of training.

Within the group of interpreters neither language proficiency (Proficiency) nor a subjective measure of the effort needed to stay at the academy (Struggle) correlated with changes in white matter microstructure.

Grey matter did not predict later language proficiency

No predictive value of cortical thickness for later proficiency was found at $p < .001$, using a clusters-extent threshold of 100 vertices. In addition, no predictive value was found for left and right hippocampal volume in relation to final proficiency.

White matter microstructure predicts later language proficiency

Within the group of interpreters, higher language proficiency was related to higher values of MD, RD and AD ($p < .05$, fully corrected for multiple comparisons) at pretest. Notably, baseline MD was higher for more proficient interpreters in areas that partly overlap with the pretest differences between interpreters and controls shown in Figure 1: forceps minor, anterior thalamic radiation, inferior fronto-occipital fasciculus, corticospinal tract, hippocampal part of the cingulum and uncinate fasciculus bilaterally. Additionally, higher MD values were found in tracts stretching along the superior longitudinal fasciculus bilaterally as well as the inferior longitudinal fasciculus in the left hemisphere (see figure 1B-D).

RD and AD showed similar trends where higher values were positively correlated with higher Proficiency. Effects for RD and AD were limited to the forceps minor bilaterally (RD) as well as parts of the inferior fronto-occipital fasciculus (RD, AD), anterior thalamic radiation (AD) and uncinate fasciculus in the left hemisphere (RD) or bilaterally (AD). No effect was found for FA. This is indicative of crossing fibers, which have been known to cause RD and AD to increase or decrease simultaneously in the same areas (Vos, Jones, Jeurissen, Viergever & Leemans, 2012). The effects in RD and AD were strongly correlated ($r(12) = .787$, $p = .001$).

The areas where MD, RD or AD correlated with language proficiency were compared to changes in grey matter from Mårtensson et al. (2012). No effects were found when controlling for multiple comparisons using Bonferroni correction.

To rule out that the correlations between DTI parameters and proficiency were driven by age differences within the group of interpreters, a bivariate correlation was performed between chronological age and the mean MD, RD and AD values from the respective skeletonized images. The significant areas from each analysis (MD, RD and AD) were then used as masks on the individual skeletonized images and the mean of all the voxels within the mask exported to SPSS for analysis. Neither MD ($r = .125$, $p = .671$), RD ($r = .237$, $p = .414$) nor AD ($r = .185$, $p = .526$) correlated with age.

The effects were not related to general intelligence

Proficiency was used once more as a covariate whilst regressing out individual scores from Raven's progressive matrices. The resulting skeletonized image was visually equal to the original analysis.

Discussion

We found that interpreters and controls differed in white matter microstructure at baseline, and that roughly the same networks were predictive of later learning proficiency in the interpreters. We did not find support for our hypothesis that white matter microstructure would change over time as an effect of training, possibly because of the prior language expertise in the interpreter population as compared to other study populations (Schlegel et al., 2012).

Of the areas differing between controls and interpreters at baseline, only the superior longitudinal fasciculus (Friederici, 2009) is strongly related to language, but it overlaps with the Arcuate fasciculus that has been tied to language performance (López-Barrosso et al., 2011, but not in Ripollés et al., 2017) and has been seen to change in response to language learning (Hosoda et al., 2013). However, the uncinate fasciculus is believed to be related to language (Catani & Mesulam, 2008), was recently implicated as relevant for language learning (Ripollés et al., 2017)

https://doi.org/10.1017/S1366728920000152 Published online by Cambridge University Press

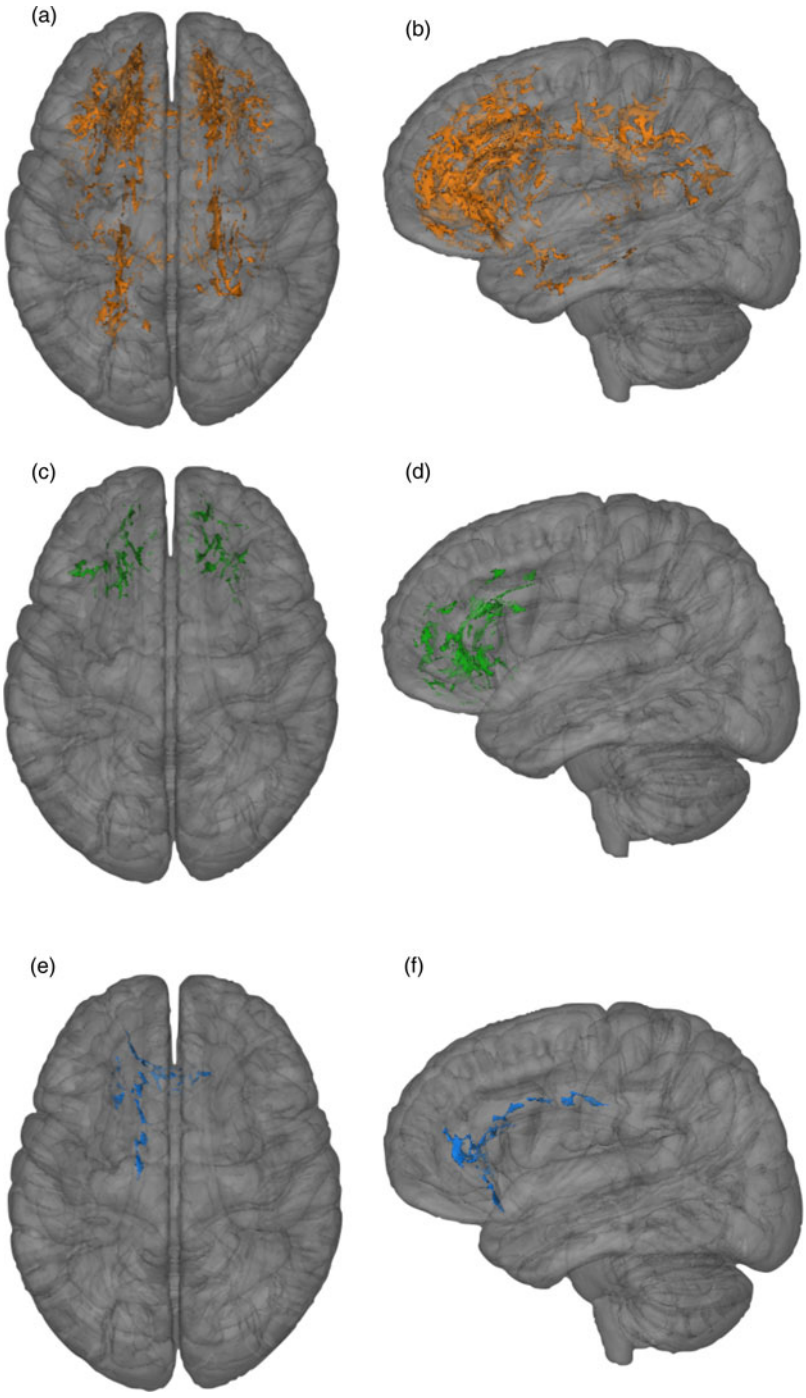


Fig. 2. (A, top; B, left) Mean diffusivity in range of areas related to language and cognitive function were positively correlated with achieved language proficiency within the interpreter group. Similar effects, but to a much lesser extent, were found for axial (C, D) and radial (E, F) diffusivity.

and belongs to the limbic system, which includes the hippocampus (Catani & Thiebaut de Schotten, 2008). The right hippocampus increased in volume in the same population as described here (Mårtensson et al., 2012). The inferior longitudinal fasciculus connects the limbic system to visual areas (Fox, Iaria & Barton, 2008) and has been observed to change in response to (Hosoda et al., 2013) and be relevant for (Ripollés et al., 2017; Qi et al., 2015), language learning in the right hemisphere. The inferior fronto-occipital fasciculus has been implied in semantic processing (Duffau, Gatignol, Mandonnet, Peruzzi, Tzourio-Mazoyer & Capelle, 2005; Duffau, Gatignol, Mortiz-Gasser & Mandonnet, 2009; Mandonnet, Nouet, Gatignol, Capelle & Duffau 2007; Ripollés et al., 2017) which is likely of relevance for interpreters with high demands on vocabulary acquisition.

Lower RD values have been connected to higher levels of myelin (Song, Sun, Ju, Lin, Cross & Neufeld, 2003; Song, Sun, Ramsbottom, Chang, Russell & Cross, 2002; Song, Yoshino, Le, Lin, Sun, Cross & Armstrong, 2005). Thus, the present findings, where controls had relatively higher values as compared to interpreters, could be taken as a cautious indication of higher degree of myelination in the group of interpreters in white-matter pathways connecting known language areas as well as areas that have been related to language learning outcomes (Schlegel et al., 2012; Mårtensson et al., 2012; Hosoda et al., 2013; López-Barrosso

https://doi.org/10.1017/S1366728920000152 Published online by Cambridge University Press

et al., 2011; Qi et al., 2015 and Ripollés et al., 2017). It should be noted, however, that these areas are far from language specific, and are used in a wide array of higher cognitive functions.

White matter microstructure (MD, RD and AD) before training was related to later language performance, which is in line with findings pointing towards the importance of brain structure at baseline in relation to learning outcomes in general (Zatorre, 2013) and for language learning specifically (Schlegel et al., 2012; Hosoda et al., 2013; López-Barrosso et al., 2011; Qi et al., 2015 and Ripollés et al., 2017). Intuitively, we would expect higher FA values to be correlated positively with performance, such as observed when Qi and colleagues (2015) measured local white matter microstructure in twenty-one English native speakers who were taught Mandarin for four weeks. Instead, we saw a relation between MD and language proficiency in the interpreters, with relatively higher MD values foreshadowing higher performance. The areas overlap with regions where interpreters differed from controls in RD, but are more widespread.

Increased MD has been observed in reduced membrane density (Sen & Basser, 2005), which should be compared to FA that is known to increase with maturation (Zatorre et al., 2012). As such our findings appear counterintuitive when compared to the supposed greater myelination of interpreters compared to controls in the same areas. Earlier training studies have shown both decrease (Taubert, Draganski, Anwander, Müller, Horstmann, Villringer & Ragert, 2010) and increase (Scholz, Klein, Behrens & Johansen-Berg, 2009) in FA values as an effect of training (Schlegel et al., 2012; Hosoda et al., 2013). Halwani, Loui, Rüber and Schlaug (2011) compared singers to instrumental musicians. Singers showed lower FA values in the arcuate fasciculus. Within the group of singers, however, the inverse was true; singers with more experience showed lower FA values. The authors of the study conclude that with more experience it is likely that microstructural complexity increases and FA values decrease.

Relatively larger values of MD, however, point towards less microstructural complexity and in the direction of large dominating fiber volumes. Since all our MD values are contained within skeletonized tracts where smaller fibers have been averaged out across participants, large MD values in themselves are not unexpected. Relatively smaller values on the other hand could point towards larger amounts of crossing fibers in less proficient interpreters. Crossing fibers have also been known to cause RD and AD to increase or decrease simultaneously in the same areas (Vos et al., 2012), which we observe in this study. This might also help explain the counterintuitive finding that RD was lower in interpreters as compared to controls, but higher in interpreters who became more proficient in the end. Perhaps there was more room for change in those interpreters, perhaps there was higher microstructural complexity. Diffusion imaging measures the effects of myelin, cell membranes, and other small structures on diffusion within relatively large voxels (2 mm) resulting in the issue of measuring microscopic anatomical factors at a macroscopic level of detail (Mori & Zhang, 2006). With the level of detail that we present in this study we cannot delve further into the neurophysiological underpinnings of our DTI findings.

It should be noted that Mårtensson et al. (2012) found selective increases in grey matter volume in the left STG, IFG, MFG and the right hippocampus for interpreters. All of these areas are adjacent to the effects observed in mean diffusivity. This makes sense, since the arcuate fasciculus is believed to connect the STG (and MTG) with the IFG and MFG (along with the premotor cortex (Catani, Jones & Ffytche, 2005; Catani, Allin, Husain, Pugliese, Mesulam, Murray & Jones, 2007). The uncinate fasciculus in turn, is believed to connect limbic areas such as the hippocampus, to frontals areas of the cortex (Catani & Thiebaut de Schotten, 2008) that correspond reasonably well with the areas observed in the current findings along with those of Schlegel and colleagues (2012). A notable difference to the grey matter findings is that the initial white matter differences between interpreters and controls are mostly bilateral whilst the gray matter changes were mainly left hemispheric (Mårtensson et al., 2012).

In contrast to some previous reports (Schlegel et al., 2012; Hosoda et al., 2013), no white matter changes were found following intensive language training. This may be due to the extensive prior experience with languages within the group of interpreters, when compared to the control group. Compared to the participants in Schlegel et al. (2012), the interpreters presumably had more extensive experience with foreign language learning. Inconsistencies could also be due to small sample size, with lower sample sizes in the range reported here and earlier (Mårtensson et al., 2012) demonstrating excessive variability when compared to larger samples (Munson & Hernandez, 2019). Hosoda et al. (2013) used a considerably larger number of participants ($n = 137$) whilst Schlegel et al. (2012) measured each participant ($n = 27$) nine times over the course of nine months.

The differences between methods used is also worth mentioning. TBSS measures changes to large tracts common to all participants, whilst Schlegel et al. (2012) measured tracts in predefined regions of interest by means of fiber-tracking as well as brain-wide analysis of white matter voxels. Training characteristics may have further contributed to the differences in findings, as the participants from Hosoda et al. (2013) knew at least 2 languages prior to training. Participants exhibited training-related changes mainly in the right hemisphere as opposed to the predominantly left hemispheric changes from Mårtensson et al. (2012), which can perhaps be taken as indication that different demands have been placed on the two populations.

We found that white matter microstructure in language-related areas predicts later language proficiency, in consistency (to some extent) with earlier findings (Schlegel et al., 2012; Hosoda et al., 2013; López-Barrosso et al., 2011; Qi et al., 2015 and Ripollés et al., 2017). Our results also fall in line with earlier cross-sectional findings showing that faster language learners have greater density of white matter in areas related to auditory processing (Golestani, Molko, Dehaene, LeBihan & Pallier, 2007) and that there are structural differences between early and late literates (Carreiras, Seghier, Baquero, Estévez, Lozano, Devlin & Price, 2009) as well as between simultaneous interpreters and controls (Elmer, Hänggi, Meyer & Jäncke, 2011).

These results are limited to a select and small group of individuals with knowledge of several foreign languages prior to training. However, they contribute to the increasing number of studies that have shown that recording white matter microstructure in the language network can be a valuable tool in understanding future learning capacity. The findings also highlight the need for further study of how initial brain structure influences and interacts with learning outcomes. Starting values need to be taken into consideration, when looking at why some individuals have an easier time learning a language than others.

Acknowledgements. The authors thank A-K. Larsson & H-O. Karlsson at UFBI as well as the staff and students at the interpreter academy of Swedish Armed Forces Intelligence and Security Center.

https://doi.org/10.1017/S1366728920000152 Published online by Cambridge University Press

This work was supported by the Sofja Kovalevskaja Award (to Martin Lövdén) from the Alexander von Humboldt foundation donated by the German Federal Ministry for Education and Research, the Swedish Research Council (421-2005-2018, 421-2010-1250), the Linnaeus environment Thinking in Time: Cognition, Communication and Learning, financed by the Swedish Research Council (349-2007-869), and grants from the Swedish Research Council and the Umeå School of Education to LN.

References

Beaulieu, C (2002). The basis of anisotropic water diffusion in the nervous system–a technical review. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, **15**, 435–455.

Bernal, B and Altman, N (2010). The connectivity of the superior longitudinal fasciculus: a tractography DTI study. *Magnetic resonance imaging*, **28**, 217–225.

Carreiras, M, Seghier, ML, Baquero, S, Estévez, A, Lozano, A, Devlin, JT and Price, CJ (2009). An anatomical signature for literacy. *Nature*, **461**, 983–986.

Catani, M and Mesulam, M (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *cortex*, **44**, 953–961.

Catani, M, Jones, DK and Ffytche, DH (2005). Perisylvian language networks of the human brain. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, **57**, 8–16.

Catani, M, Allin, MP, Husain, M, Pugliese, L, Mesulam, MM, Murray, RM and Jones, DK (2007). Symmetries in human brain language pathways correlate with verbal recall. *Proceedings of the National Academy of Sciences*, **104**, 17163–17168.

Catani, M and Thiebaut de Schotten, M (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, **44**, 1105–1132.

Dahlquist, H (2004). Vem klarar tolkskolan? *Pliktverkets rapportserie*.

Davis, MH and Gaskell, MG (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 3773–3800.

Dale, AM and Sereno, MI (1993). Improved localizadon of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *Journal of cognitive neuroscience*, **5**, 162–176.

Dale, AM, Fischl, B and Sereno, MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194.

Desikan, RS, Segonne, F, Fischl, B, Quinn, BT, Dickerson, BC, Blacker, D, Buckner, RL, Dale, AM, Maguire, RP, Hyman, BT, Albert, MS and Killiany, RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980.

Demonet, JF (2005). Renewal of the Neurophysiology of Language: Functional Neuroimaging. *Physiological Reviews*, **85**, 49–95.

Duffau, H, Gatignol, P, Mandonnet, E, Peruzzi, P, Tzourio-Mazoyer, N and Capelle, L (2005) New insights into the anatomo-functional connectivity of the semantic system: a study using cortico-subcortical electrostimulations. *Brain*, **128**:797–810.

Duffau, H, Gatignol, P, Moritz-Gasser, S and Mandonnet, E (2009) Is the left uncinate fasciculus essential for language? A cerebral stimulation study. *Journal of Neurology*, **256**:382–389.

Elmer, S, Hänggi, J, Meyer, M and Jäncke, L (2011). Differential language expertise related to white matter architecture in regions subserving sensorymotor coupling, articulation, and interhemispheric transfer. *Human Brain Mapping*, **32**, 2064–2074.

Filaire, E, Sagnol, M, Ferrand, C, Maso, F and Lac, G (2001). Psychophysiological stress in judo athletes during competitions. *The Journal of sports medicine and physical fitness*, **41**, 263.

Fox, CJ, Iaria, G and Barton, JJS (2008). Disconnection in prosopagnosia and face processing. *Cortex*, **44**, 996–1009.

Fischl, B and Dale, AM (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, **97**, 11050–11055.

Fischl, B, Liu, A and Dale, AM (2001). Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE transactions on medical imaging*, **20**, 70–80.

Fischl, B, Salat, DH, Busa, E, Albert, M, Dieterich, M, Haselgrove, C, van der Kouwe, A, Killiany, R, Kennedy, D, Klaveness, S, Montillo, A, Makris, N, Rosen, B and Dale, AM (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355.

Fischl, B, Salat, DH, Van Der Kouwe, AJ, Makris, N, Ségonne, F, Quinn, BT and Dale, AM (2004). Sequence-independent segmentation of magnetic resonance images. *Neuroimage*, **23**, S69–S84.

Fischl, B, Sereno, MI and Dale, AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195–207.

Fischl, B, Sereno, MI, Tootell, RB and Dale, AM (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping* **8**, 272–284.

Fischl, B, van der Kouwe, A, Destrieux, C, Halgren, E, Segonne, F, Salat, DH, Busa, E, Seidman, LJ, Goldstein, J, Kennedy, D, Caviness, V, Makris, N, Rosen, B and Dale, AM (2004) Automatically parcellating the human cerebral cortex. *Cerebral Cortex* **14**, 11–22.

Friederici, AD (2009). Pathways to language: fiber tracts in the human brain. *Trends in cognitive sciences*, **13**, 175–181.

Friederici, AD (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, **91**, 1357–1392.

Friederici, AD and Gierhan, SM (2013). The language network. *Current Opinion in Neurobiology*, **23**, 250–254.

Golestani, N, Molko, N, Dehaene, S, LeBihan, D and Pallier, C (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, **17**, 575–582.

Halwani, GF, Loui, P, Rüber, T and Schlaug, G (2011). Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians. *Frontiers in Psychology*, **2**.

Han, X, Jovicich, J, Salat, D, van der Kouwe, A, Quinn, B, Czanner, S, Busa, E, Pacheco, J, Albert, M, Killiany, R, Maguire, P, Rosas, D, Makris, N, Dale, A, Dickerson, B and Fischl, B (2006) Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage* **32**, 180–194.

Hickok, G and Poeppel, D (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, **8**, 393–402.

Hosoda, C, Tanaka, K, Nariai, T, Honda, M and Hanakawa, T (2013). Dynamic neural network reorganization associated with second language vocabulary acquisition: A multimodal imaging study. *Journal of Neuroscience*, **33**, 13663–13672.

Jenkinson, M and Smith, S (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis*, **5**, 143–156.

Jenkinson, M, Bannister, P, Brady, M and Smith, S (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, **17**, 825–841.

Jovicich, J, Czanner, S, Greve, D, Haley, E, van der Kouwe, A, Gollub, R, Kennedy, D, Schmitt, F, Brown, G, Macfall, J, Fischl, B and Dale, A (2006). Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage* **30**, 436–443.

Kassem, MS, Lagopoulos, J, Stait-Gardner, T, Price, WS, Chohan, TW, Arnold, J, Hatton, SN and Bennett, MR (2012). Stress-Induced Grey Matter Loss Determined by MRI Is Primarily Due to Loss of Dendrites and Their Synapses. *Molecular Neurobiology*, **47**, 645–661.

Li, P, Legault, J and Litcofsky, KA (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. *Cortex*, **58**, 301–324.

Lopez-Barroso, D, de Diego-Balaguer, R, Cunillera, T, Camara, E, Münte, TF and Rodriguez-Fornells, A (2011). Language learning under working memory constraints correlates with microstructural differences in the ventral language pathway. *Cerebral cortex*, **21**, 2742–2750.

Mandonnet, E, Nouet, A, Gatignol, P, Capelle, L and Duffau, H (2007) Does the left inferior longitudinal fasciculus play a role in language? A brain stimulation study. *Brain* **130**:623–629.

https://doi.org/10.1017/S1366728920000152 Published online by Cambridge University Press

Mori, S and Zhang, J (2006). Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron*, **51**, 527–539.

Munson, BA and Hernandez, AE (2019). Inconsistency of findings due to low power: A structural MRI study of bilingualism. *Brain and language*, **195**, 104642.

Mårtensson, J, Eriksson, J, Bodammer, NC, Lindgren, M, Johansson, M, Nyberg, L and Lövdén, M (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, **63**, 240–244.

Paulesu, E, Vallar, G, Berlingeri, M, Signorini, M, Vitali, P, Burani, C, Perani, D and Fazio, F (2009). Supercalifragilisticexpialidocious: how the brain learns words never heard before. *Neuroimage*, **45**, 1368–1377.

Pliatsikas, C, Moschopoulou, E and Saddy, JD (2015). The effects of bilingualism on the white matter structure of the brain. *Proceedings of the National Academy of Sciences*, **112**, 1334–1337.

Price, CJ (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, **1191**, 62–88.

Qi, Z, Han, M, Garel, K, Chen, ES and Gabrieli, JDE (2015). White-matter structure in the right hemisphere predicts Mandarin Chinese learning success. *Journal of Neurolinguistics*, 1–15.

Ripollés, P, Biel, D, Peñaloza, C, Kaufmann, J, Marco-Pallarés, J, Noesselt, T and Rodríguez-Fornells, A (2017). Strength of temporal white matter pathways predicts semantic learning. *Journal of Neuroscience*, 1720–17.

Sakai, KL (2005). Language acquisition and brain development. *Science*, **310**, 815–819.

Ségonne, F, Dale, AM, Busa, E, Glessner, M, Salat, D, Hahn, HK and Fischl, B (2004). A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, **22**, 1060–1075.

Schlegel, AA, Rudelson, JJ and Tse, PU (2012). White matter structure changes as adults learn a second language. *Journal of cognitive neuroscience*, **24**, 1664–1670.

Scholz, J, Klein, MC, Behrens, TEJ and Johansen-Berg, H (2009). Training induces changes in white-matter architecture. *Nature neuroscience*, **12**, 1370–1371.

Sen, PN and Basser, PJ (2005). A Model for Diffusion in White Matter in the Brain. *Biophysical Journal*, **89**, 2927–2938.

Sheppard, JP, Wang, JP and Wong, PC (2012). Large-scale cortical network properties predict future sound-to-word learning success. *Journal of Cognitive Neuroscience*, **24**, 1087–1103.

Shtyrov, Y (2011). Fast mapping of novel word forms traced neurophysiologically. *Frontiers in psychology*, **2**, 340.

Shtyrov, Y, Nikulin, VV and Pulvermüller, F (2010). Rapid Cortical Plasticity Underlying Novel Word Learning. *The Journal of Neuroscience*, **30**, 16864–16867. http://doi.org/10.1523/JNEUROSCI.1376-10.2010

Smith, SM, Jenkinson, M, Johansen-Berg, H, Rueckert, D, Nichols, TE, Mackay, CE, Watkins, KE, Ciccarelli, O, Cader, MZ, Matthews, PM and Behrens, TE (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage*, **31**, 1487–1505.

Smith, SM, Jenkinson, M, Woolrich, MW, Beckmann, CF, Behrens, TE, Johansen-Berg, H, Bannister, PR, De Luca, M, Drobnjak, I, Flitney, DE and Niazy, RK (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, **23**, pp.S208–S219.

Song, SK, Sun, SW, Ju, WK, Lin, SJ, Cross, AH and Neufeld, AH (2003). Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. *Neuroimage*, **20**, 1714–1722.

Song, SK, Sun, SW, Ramsbottom, MJ, Chang, C, Russell, J and Cross, AH (2002). Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. *Neuroimage*, **17**, 1429–1436.

Song, SK, Yoshino, J, Le, TQ, Lin, SJ, Sun, SW, Cross, AH and Armstrong, RC (2005). Demyelination increases radial diffusivity in corpus callosum of mouse brain. *Neuroimage*, **26**, 132–140.

Spielberger, CD, Gorsuch, RL and Lushene, RE (1970). Stai. *Manual for the State-Trait Anxiety Inventory (Self Evaluation Questionnaire). Palo Alto California: Consulting Psychologist*, **22**, 1–24.

Taubert, M, Draganski, B, Anwander, A, Müller, K, Horstmann, A, Villringer, A and Ragert, P (2010). Dynamic properties of human brain structure: learning-related changes in cortical areas and associated fiber connections. *The Journal of Neuroscience*, **30**, 11670–11677.

Ullén, F, Hambrick, DZ and Mosing, MA (2016). Rethinking expertise: A multifactorial gene–environment interaction model of expert performance. *Psychological Bulletin*, **142**, 427–446.

Vos, SB, Jones, DK, Jeurissen, B, Viergever, MA and Leemans, A (2012). The influence of complex white matter architecture on the mean diffusivity in diffusion tensor MRI of the human brain. *Neuroimage*, **59**, 2208–2216.

Xiang, H, Dediu, D, Roberts, L, Oort, EV, Norris, DG and Hagoort, P (2012). The structural connectivity underpinning language aptitude, working memory, and IQ in the perisylvian language network. *Language learning*, **62**, 110–130.

Yang, J, Gates, KM, Molenaar, P and Li, P (2015). Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics*, **33**, 29–49.

Zatorre, RJ (2013). Predispositions and Plasticity in Music and Speech Learning: Neural Correlates and Implications. *Science*, **342**, 585–589.

Zatorre, RJ, Fields, RD and Johansen-Berg, H (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature neuroscience*, **15**, 528–536.

https://doi.org/10.1017/S1366728920000152 Published online by Cambridge University Press