


## RESEARCH ARTICLE

# Cortical and white matter correlates of language-learning aptitudes

Mikael Novén<sup>1</sup>  | Hampus Olsson<sup>2</sup> | Gunther Helms<sup>2</sup> | Merle Horne<sup>1</sup> | Markus Nilsson<sup>3</sup> | Mikael Roll<sup>1</sup>

<sup>1</sup>Department of Linguistics and Phonetics, Lund University, Lund, Sweden

<sup>2</sup>Department of Clinical Sciences Lund, Medical Radiation Physics, Lund University, Lund, Sweden

<sup>3</sup>Department of Clinical Sciences Lund, Radiology, Lund University, Lund, Sweden

## Correspondence

Mikael Novén, Department of Linguistics and Phonetics, Lund University, Språk-och Litteraturcentrum, Box 201, Lund 221 00, Sweden.

Email: mikael.noven@ling.lu.se

## Funding information

the Swedish Research Council, Grant/Award Number: 2018.00632; Marcus and Amalia Wallenberg Foundation, Grant/Award Number: 2018.0021; Knut and Alice Wallenberg Foundation, Grant/Award Numbers: 2018.0454, 2014.0139

## Abstract

People learn new languages with varying degrees of success but what are the neuroanatomical correlates of the difference in language-learning aptitude? In this study, we set out to investigate how differences in cortical morphology and white matter microstructure correlate with aptitudes for vocabulary learning, phonetic memory, and grammatical inferring as measured by the first-language neutral LLAMA test battery. We used ultra-high field (7T) magnetic resonance imaging to estimate the cortical thickness and surface area from sub-millimeter resolved image volumes. Further, diffusion kurtosis imaging was used to map diffusion properties related to the tissue microstructure from known language-related white matter tracts. We found a correlation between cortical surface area in the left posterior-inferior precuneus and vocabulary learning aptitude, possibly indicating a greater predisposition for storing word-figure associations. Moreover, we report negative correlations between scores for phonetic memory and axial kurtosis in left arcuate fasciculus as well as mean kurtosis, axial kurtosis, and radial kurtosis of the left superior longitudinal fasciculus III, which are tracts connecting cortical areas important for phonological working memory.

## KEYWORDS

cortical morphometry, cortical surface area, cortical thickness, dwMRI, language-learning aptitude, ultra-high field

## 1 | INTRODUCTION

The study of what makes a good language learner has been a growing research field since the 1950's (Carroll & Sapon, 1959). The need and benefit for people to learn new languages far into adulthood has only grown with globalization, which further spurs the interest to understand the nature of language-learning aptitude. Language-learning aptitude is a stable trait within individuals (Granena, 2013) that predicts how well people can learn a foreign language (Li, 2014). Brain structure has been shown to correlate with various aspects of language-related performance, ranging from low-level acoustic

processing to executive control of languages in terms of fluency and speech-in-noise processing (Golestani, 2012). Heschl's gyrus (HG) even differs in shape between expert phoneticians and nonexperts, probably already before phonetic training (Golestani, Price, & Scott, 2011). Research into the associations between language-learning aptitude and cortical morphology has found that the grey matter volume or shape of HG affects the ability to learn Mandarin word tones (Wong et al., 2008), performance on a speech imitation task including novel phonological contrasts (Turker, Reiterer, Seither-Preisler, & Schneider, 2017), and general language-learning aptitude (Turker, Reiterer, Schneider, & Seither-Preisler, 2019). In addition,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

cortical thickness of anterior insula correlates with the aptitude for learning to discriminate new speech sounds in bilinguals but not monolinguals (Rodríguez, Archila-Suerte, Vaughn, Chiarello, & Hernandez, 2018). Furthermore, Novén, Schremm, Nilsson, Horne, and Roll (2019) found the cortical thickness of Brodmann area (BA) 45 in the left inferior frontal gyrus (IFG) to correlate with grammatical inferencing ability. Studies including measures of white matter microstructure have instead observed associations between artificial grammar learning ability and the structure of connections to the left IFG (Flöel, de Vries, Scholz, Breitenstein, & Johansen-Berg, 2009) and structure of right-hemispheric white matter and successful learning of Mandarin (Qi, Han, Garel, San Chen, & Gabrieli, 2015). Moreover, white matter microstructure in frontal networks predicts language learning proficiency in conscript interpreters (Mårtensson et al., 2020). However, previous studies have not specifically related to language-learning aptitude theory, which strives toward understanding what cognitive processes are involved in making a good language learner (Wen, Biedroń, & Skehan, 2016). Thus, language-learning aptitude has most often been inferred by learning outcomes in a classroom environment (Mårtensson et al., 2020; Qi et al., 2015) or in learning isolated language elements (Flöel et al., 2009; Rodríguez et al., 2018; Turker et al., 2017; Wong et al., 2008). In fact, only two studies we know of, apart from our previous work (Novén et al., 2019), have examined brain structural correlates of behavioral measures rooted in language-learning aptitude theory (Turker et al., 2019; Xiang et al., 2012). With regards to the parameters of brain structure studied, some have only investigated cortical morphology in predefined regions of interest (ROIs; Rodríguez et al., 2018; Turker et al., 2017, 2019; Wong et al., 2008), while others have been limited in spatial resolution (Novén et al., 2019). Further, cortical surface area has been overlooked as a potential correlate to language-learning aptitude. This is despite the fact that cortical thickness and cortical surface area are independent measures of cortical morphology (Meyer, Liem, Hirsiger, Jäncke, & Hänggi, 2013; Panizzon et al., 2009; Vuoksimaa et al., 2015). Also, only standard diffusion tensor imaging (DTI) has previously been used to examine tissue microstructure correlates of language-learning aptitude (Flöel et al., 2009; Mårtensson et al., 2020; Qi et al., 2015; Xiang et al., 2012). Using instead diffusion kurtosis imaging (DKI), it is possible to capture more detailed information about the tissue microstructure (Jensen & Helpert, 2010). Building on the work in Novén et al. (2019), the present study investigates the cortical morphometry, that is, the thickness and surface area of the cortex, at sub-millimeter resolution. Further, DKI is used to inspect the microstructure of language-relevant white matter tracts. Language-learning aptitude is measured by the LLAMA test battery (Meara, 2005), based on the work of Carroll (1962). The refined techniques allow for measuring anatomical correlates of language-learning aptitude in more detail than previous studies.

## 1.1 | LLAMA test battery and brain structure

Language-learning aptitude can be assessed using the first language-independent, computer-based LLAMA test battery (Meara, 2005). It

has four sub-tests assessing different aptitude components: written vocabulary (LLAMA B), phonetic memory (LLAMA D), sound-symbol correspondence (LLAMA E), and grammatical inferencing (LLAMA F). The brain structure correlates of LLAMA test scores have been investigated in three studies, described below. Grey matter volume in the right, but not left, HG correlates with LLAMA scores from all sub-tests except phonetic memory in children, 10–16 years of age (Turker et al., 2019). Sound-symbol correspondence aptitude has been shown to correlate with the fractional anisotropy (FA) in a tract connecting BA45 in each hemisphere (Xiang et al., 2012). Moreover, vocabulary learning aptitude has been observed to correlate with a difference between FA in left and right tracts between BA47 and the parietal lobe (Xiang et al., 2012). We have previously seen a correlation between cortical thickness in left BA45 and medial frontal gyrus and grammatical inferencing aptitude (Novén et al., 2019). This is in line with the finding that the left-lateralization of FA in the connection between BA 45 and the posterior temporal lobe as well as the sum of the number of streamlines between BA6 and the posterior temporal lobe from both hemispheres correlates with grammatical inferencing aptitude (Xiang et al., 2012). Taken together, previous studies have reported correlates between LLAMA test scores and cortical volume in a predefined ROI (Turker et al., 2019) or using data of lower resolution for cortical thickness estimation (Novén et al., 2019) and no studies have investigated possible correlations with cortical surface area. Tissue microstructure in white matter tracts is associated with language-learning aptitude but has only been investigated using standard DTI (F. C. K. Wong, Chandrasekaran, Garibaldi, & Wong, 2011; Xiang et al., 2012) and construction of white matter tracts through probabilistic tracking from seed regions derived from functional experiments (Xiang et al., 2012). This method runs the risk of including nonanatomical white matter tracts (Schilling et al., 2019).

## 1.2 | White matter tract segmentation

Four anatomical white matter tracts are essential for language processing and could be relevant for language-learning aptitudes: the arcuate fasciculus (AF), subcomponent three of the superior longitudinal fasciculus (SLF III), the uncinate fasciculus (UF), and the inferior frontal-occipital fasciculus (IFOF). AF connects the IFG and the middle frontal gyrus with the posterior superior temporal gyrus as well as the temporal occipital transition region (Catani, Jones, & ffytche, 2005; Makris et al., 2004). The left AF is used for mapping sound features to articulatory representations (Saur et al., 2008; Wong et al., 2011) and complex syntactic processing (Friederici & Gierhan, 2013). SLF links the frontal lobe with the lateral occipital and temporal lobes and is divided into three subcomponents, I–III (Makris et al., 2004). SLF III connects the supramarginal gyrus with the prefrontal and ventral premotor cortices and is the SLF subcomponent most implicated in neurolinguistic research due to its involvement in the ability to repeat speech (Friederici & Gierhan, 2013; Saur et al., 2008). UF joins the anterior temporal lobe and the frontal lobe (Catani & Thiebaut de Schotten, 2008). IFOF extends between the occipital lobe and the orbito- and inferior frontal cortices (Catani & Thiebaut de Schotten, 2008). Both UF and

IFOF are suggested to be important for being able to map sounds to meaning (Wong et al., 2011). UF is further involved in the construction of short syntactic phrases (Friederici & Gierhan, 2013).

From diffusion-weighted MRI data, it is possible to compute parameters that describe white matter tissue microstructure. Detailed descriptions and derivations of the parameters can be found in for example, Pierpaoli, Jezzard, Basser, Barnett, and Di Chiò (1996) and Jensen and Helpem (2010) for DTI and DKI, respectively. Put simply, spontaneous movement of water molecules within tissues occurs at different rates (diffusivities) along different spatial directions, depending on the microstructure of the tissue. DTI yields measures of the diffusivities along different directions, and thus an understanding of the anisotropy (i.e., how much easier it is for water to move along the most unrestricted direction than along the most restricted direction) of the tissue. However, DTI is based on the assumption that the diffusion properties are Gaussian, meaning that, along a given direction, the tissue is assumed to be homogenous (i.e., no variability in the obstruction of the water movement). Important DTI parameters are the mean diffusivity (MD), the mean apparent diffusivity across all diffusion encoding angles, and axial and radial diffusivity (AD and RD), the diffusivities along and perpendicular to the direction of greatest diffusivity. FA quantifies how much greater AD is than RD and ranges from 0 (water can move as easily in all directions) to 1 (water can only move along one direction). However, in complex tissues and macromolecules restrict diffusion, the water displacement profile becomes less Gaussian. The shape of the displacement profile is captured by the kurtosis (the standardized and normalized fourth central moment of the displacement probability distribution function). Axial and radial kurtoses (AK and RK) describe the diffusion kurtoses along or perpendicular to the principal diffusion direction while mean kurtosis (MK) is the mean kurtosis across all encoding directions. The diffusivity parameters (MD, AD, RD, and FA) are thus measures of the voxel-average diffusion profile while the diffusion kurtosis parameters (MK, AK, and RK) give an estimation of the sub-voxel structure of the tissue. It is also possible to use diffusion-weighted MRI to construct white matter tracts through probabilistic tracking from seed regions derived from functional experiments (Xiang et al., 2012), but this is at the risk of including nonanatomical, that is, false, white matter tracts (Schilling et al., 2019). Therefore, we used a white matter segmentation software tool (TractSeg) to automatically segment anatomically relevant, language-related tracts (Wasserthal, Neher, Hirjak, & Maier-Hein, 2019; Wasserthal, Neher, & Maier-Hein, 2018a, 2018b).

Taken together, the tissue microstructure of the language-related tracts could be a source for or reflection of language-learning aptitude as the tracts structurally connect, that is, allow for signal transport between cortical areas important for language processing. Therefore, we extracted mean diffusion parameter values, reflecting the microstructure of the tissue, from mentioned tracts and tested their correlations with LLAMA test scores.

### 1.3 | The present study

The present study investigated the neuroanatomical correlates of language-learning aptitude using the LLAMA tests and ultra-high field

MRI. The benefits of using ultra-high field MRI lie in the increased signal-to-noise ratio and consecutive tissue contrast (Duyn, 2012) allowing for high resolutions, limiting partial volume effects, and yielding more effective tissue segmentations from T1-weighted image volumes (Zaretskaya, Fischl, Reuter, Renvall, & Polimeni, 2018). This study contributes to the understanding of the structural neural correlates of language-learning aptitude by using higher spatial resolution than in previous studies (Wong et al., 2011; Xiang et al., 2012), as a basis for cortical morphometry. Furthermore, utilizing DKI allows for the calculation of more detailed tissue microstructure parameters of relevant white matter tracts as compared to standard DTI.

## 2 | METHODS

### 2.1 | Participants

Fifty-seven university students (15 men, 42 women) growing up in monolingual families (35 German and 22 Swedish) were recruited for this study. Mean age was 22.7 years (range = 20–27 years). No participant had any history of psychiatric disorders. All participants were right-handed as defined as a minimum of +25 in the Edinburgh handedness index (Oldfield, 1971) and had vision that was normal or corrected to normal. To ensure normal hearing (necessary for the phonetic memory subtest), a minimal hearing threshold of <20 dB for pure tones of 250, 500, 1,000, 2,000, 4,000, and 8,000 Hz frequency was required to participate. Hearing thresholds were measured using fixed-frequency Békésy audiometry in the same way and using the same equipment as in Novén et al. (2019).

Participants were characterized concerning their fluid intelligence, working memory capacity, and musical sophistication. Fluid intelligence was assessed by a short-form of the Raven's matrices (Raven, 2000) described in Mårtensson and Lövdén (2011). Participants were given 10 min to complete 18 matrices by selecting the missing ninth pattern based on eight given patterns. Scores were equal to the number of correctly chosen patterns. Participants' working memory capacity was measured using an automated version of the operation span test (Unsworth, Heitz, Schrock, & Engle, 2005). Participants were required to solve arithmetic problems while remembering series of letters. The final score is the total number of correctly recalled letters. Musical sophistication was judged using the Goldsmith musical sophistication index (Gold-MSI; Müllensiefen, Gingras, Musil, & Stewart, 2014). The local ethics board approved the study and all participants gave written consent prior to the experiment.

### 2.2 | LLAMA tests

To measure participants' language-learning aptitude, three of the four LLAMA tests were administered: The vocabulary (LLAMA B) subtest, the phonetic memory (LLAMA D) subtest, and the grammatical inferencing (LLAMA F) subtest. The fourth subtest, LLAMA E, focusing on sound-symbol correspondence, was left out.

In the LLAMA B subtest, participants were shown 20 cartoon objects and were given the written name of each object when clicking on them with the computer mouse. Each participant was given 2 min to learn as many names as possible. In the test phase, each name was given and the participant was instructed to click the corresponding object on the screen. Feedback was given on the validity of the participant's choice for each answer. LLAMA B can thus be considered to test proficiency in identifying written word-picture meaning correspondences.

The LLAMA D subtest started with the participants being asked to listen to a short set of spoken foreign words. Immediately afterward, either words found in the initial phase of the test or new words were played to the participant who was asked to indicate if the word was part of the initial set of words or not. This test can be assumed to assess phonetic memory capacity.

The LLAMA F subtest consists of 20 image-sentence pairs consisting of pictures of stylized figures performing different actions. The images are associated with sentences consisting of strings of written pseudowords. The pictures and word strings were shown to participants as they clicked on boxes on the screen in a 5-min training phase. They were told that they should use the time to learn as much as possible about the language used in the sentences to describe the images. Taking written notes was allowed for this subtest but not for the others. In the test phase, the participants were told to choose between two (one correct and one incorrect) sentence describing a new or old picture.

The scores for each LLAMA subtest range from 0 to 100.

LLAMA test scores have been found to correlate well with learning various aspects of a second language (Abrahamsson & Hyltenstam, 2008; Granena, 2012, 2013; Granena & Long, 2013). The tests have been shown to be internally consistent and test scores are stable over time within participants (Granena, 2013). The scores are unaffected by gender and age for test-takers above 12 years of age but level of formal education is significantly positively correlated with LLAMA B, E, and F but not D scores (Rogers et al., 2016; Rogers, Meara, Barnett-Legh, Curry, & Davie, 2017). Hence, we take the LLAMA test battery to be accurate and stable measures of language-learning aptitude components.

## 2.3 | MRI acquisition

MRI was performed with an actively shielded 7T scanner (Achieva, Philips, Best, Netherlands) equipped with a two-channel transmit and 32 channel receive phased-array head coil (Nova Medical, Wilmington, MA). Radiofrequency (B1) field inhomogeneities were reduced by using dielectric pads (Teeuwisse, Brink, & Webb, 2012). A T1-weighted 3D magnetization-prepared rapid gradient echo (MPRAGE) sequence with repetition time (TR) = 8 ms, echo time (TE) = 1.97 ms, flip angle = 8°, inversion time (TI) = 1,200 ms, outer-loop SENSE-factor = 2.5,  $0.8 \times 0.8 \times 0.8 \text{ mm}^3$  voxels with a scan time of 5:13 min was acquired. For normalization, a proton density (PD)-weighted gradient echo sequence at identical TR and TE = 1.97

but lower flip angle (2°) and resolution ( $1.6 \times 1.6 \times 1.6 \text{ mm}^3$ ) with a scan time of 54 s was also acquired (Olsson & Helms, 2021).

Two diffusion-weighted image volumes were acquired. One single-shelled with a higher angular resolution, henceforth called the tractography volume, and one multi-shelled, hereafter the DKI volume, to permit for DKI estimation. The tractography volume was acquired with TR = 9.6 s, TE = 73 ms, flip angle = 90°, SENSE factor = 1.5, partial Fourier = 0.642,  $2 \times 2 \times 2 \text{ mm}^3$  voxels, 56 directions and  $b = [0 \ 2,000] \text{ s/mm}^2$  giving a scan time of 9:27 min. Additionally, one extra  $b = 0$  volume with a flipped phase encoding direction was acquired for correcting susceptibility induced distortions. The DKI volume was acquired with TR = 6.5 s, TE = 88 ms, flip angle = 90°, SENSE factor = 2, partial Fourier = 0.75,  $2 \times 2 \times 4 \text{ mm}^3$  voxels,  $b = [0 \ 100 \ 700 \ 1,400 \ 2,000] \text{ s/mm}^2$  and  $[1 \ 6 \ 6 \ 12 \ 16]$  encoding directions giving a scan time of 4:40 min. The ordering of the encoding strengths were randomized in order to temporally spread the energy consumption and avoid temporal confounds (Hutter et al., 2018; Vos et al., 2017).

The collected MRI data is available in OpenNeuro Dataset ds003508 (Novén et al., 2021).

## 2.4 | Cortical morphometry

For each participant, the PD-weighted reference volume was rigidly registered to the T1-weighted volume using the FMRIB's Linear Image Registration Tool (Jenkinson, Bannister, Brady, & Smith, 2002; Jenkinson & Smith, 2001). Subsequently, the T1-weighted image was divided by the PD-weighted reference to mitigate B1 inhomogeneities and eliminate the influence of PD-weighting and effective transverse ( $T2^*$ ) relaxation. This approach is based on the concept described in Van de Moortele et al. (2009). These volumes were brain extracted using the Brain Extraction Tool (BET) in FSL (Smith, 2002), quality assured and manually edited when needed to ensure that as little non-brain tissue as possible was included in the segmentation. The normalized and brain-extracted T1-weighted image volumes then underwent nonparametric nonuniform bias field correction, to remove residual influence of B1 inhomogeneities (version 2.1.0.post685-g86e09; Tustison et al., 2010).

Cortical reconstruction and volumetric segmentation was performed with the FreeSurfer image analysis suite (v 6.0), which is documented and freely available for download online (<http://surfer.nmr.mgh.harvard.edu/>). The technical details of the analysis pipelines and tools for generating aligned maps of cortical thickness and surface area measures have been described elsewhere (Dale, Fischl, & Sereno, 1999; Desikan et al., 2006; Destrieux, Fischl, Dale, & Halgren, 2010; Fischl, Sereno, & Dale, 1999; Fischl et al., 2004; Greve & Fischl, 2009) but a brief summary of and departures from standard analyses follows below.

The standard recon-all pipeline was run with the `-hires` flag (Zaretskaya et al., 2018), except for the skull-stripping performed as described above. The processing pipeline included motion-correction, intensity normalization, and tessellation and refinement of the white/

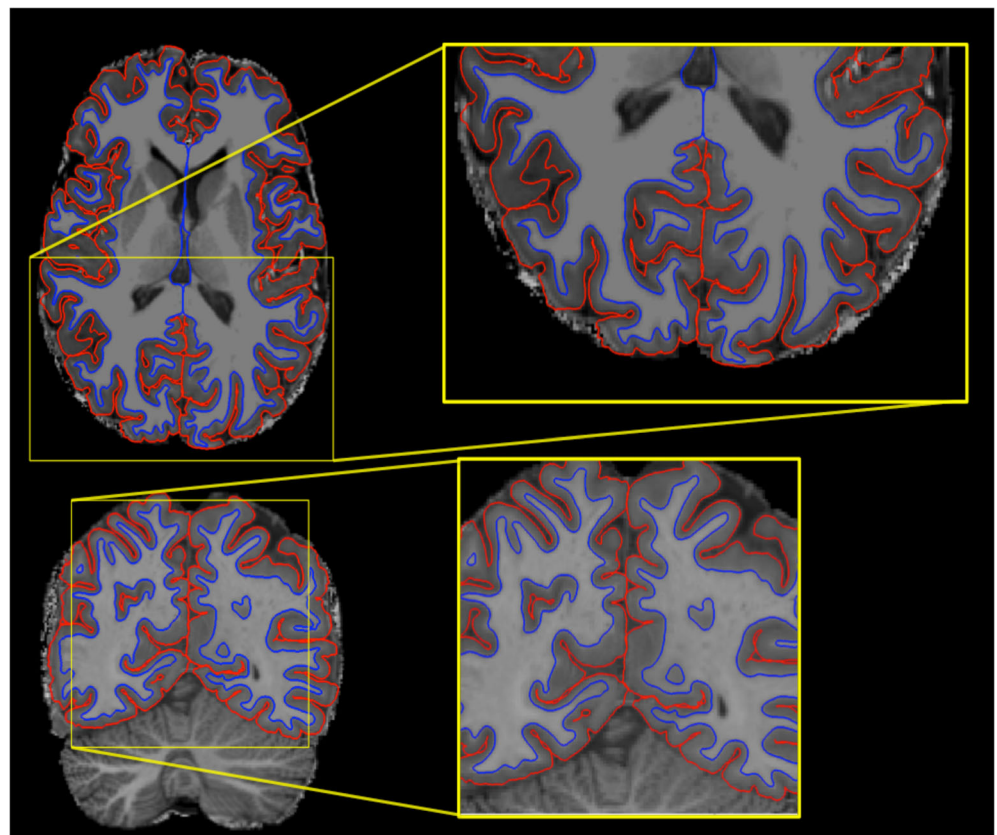
grey matter border, from now on called the white surface, and grey/cerebrospinal fluid border, henceforth the pial surface. Cortical surface area is calculated as the area of the tessellation triangles and mapped onto the vertices as the mean of the triangles the vertex is part of. Cortical thickness was computed as the shortest distance between points on the white surface to the pial surface. Maps of cortical thickness and surface area were mapped onto the FreeSurfer common space, *FsAverage*, and smoothed using a 15 mm full width at half maximum (FWHM) Gaussian kernel. All segmentations and surface reconstructions were visually inspected and manually corrected as needed but the output from one participant was deemed too poor and was left out of further analysis. An example of the resulting surface reconstructions is shown in Figure 1.

## 2.5 | DWI processing

The flipped phase encoding volume was used to correct the tractography volume from susceptibility-induced errors using the *topup* tool in FSL (Andersson, Skare, & Ashburner, 2003; Smith et al., 2004). At this stage, the tractography volume was corrected for eddy currents and subject motion using *ElastiX* with extrapolated target volumes (Nilsson, Szczepankiewicz, van Westen, & Hansson, 2015). A first estimation of FA was taken from the *TractSeg* tool *calc\_FA* and the parameter map was used to register the tractography volume to MNI space using an FA-template included in

FSL. White matter bundle segmentation was performed on the preprocessed tractography volume using the *TractSeg* tool openly available at <https://github.com/MIC-DKFZ/TractSeg> (Wasserthal et al., 2018a, 2018b; Wasserthal et al., 2019). *TractSeg* uses a fully convolutional neural network trained on a cohort from the Human Connectome Project (Van Essen et al., 2013) to automatically segment well-known anatomical tracts. This is done based on fiber orientation distribution functions obtained from constrained spherical deconvolution (Tournier, Calamante, & Connelly, 2007). The chosen tract segmentations were manually inspected to ensure reasonable segmentation performance. The tractography volume could not be acquired from two participants due to problems with the scanner and data from one participant was discarded from further analysis due to signal loss in the tractography volume in temporal and inferofrontal areas impacting the quality of the segmentations.

The DKI volume underwent correction for noise, Gibbs ringing, Rician biases, and signal outliers followed by motion and eddy current correction identically as for the tractography volume, described above. Thereafter, the DKI volume was smoothed using a 2.4 mm FWHM Gaussian kernel and MD, FA, AD, RD, MK, AK, and RK were estimated using the freely available *DESIGNER* tool (<https://github.com/NYU-DiffusionMRI/DESIGNER>; Ades-Aron et al., 2018) and registered to MNI space. The mean parameter values of each tract were then extracted using the *Tractometry* command in *TractSeg*. Parameters from IFOF were only obtainable from 10 of the subjects. Therefore, IFOF was not included in the analyses.



**FIGURE 1** Example of the resulting FreeSurfer surface reconstructions. Pial surfaces are shown in red and white surfaces in blue



## 2.6 | Statistical analyses

To assess the risk that differences in LLAMA test scores reflect differences in other cognitive capacities or linguistic background, we investigated correlations between working memory, fluid intelligence, musicality, or number of learned languages and LLAMA test scores. Such correlations would limit the interpretation of correlations between LLAMA scores and cortical morphometry or tissue microstructure. Moreover, the LLAMA subtests are supposed to measure independent components of language-learning aptitude and this should be verified. Therefore, the interdependence of LLAMA scores as well as correlations with working memory, general intelligence, musicality, and number of learned languages was assessed using Spearman's rank correlation analyses.

Age and gender were covariates of no interest in all cortical and diffusion parameter analyses. The FreeSurfer tool *qdec* was used for correlation analyses of cortical thickness and surface area measures. Brain size correlates with cortical thickness and surface area (Im et al., 2008). Therefore, estimated intracranial volume (eICV), as given by FreeSurfer, was included as a covariate of no interest in all cortical thickness and surface area correlation tests. Clusters in which cortical thickness or surface area correlated with a LLAMA test score underwent comparison to cluster size limits derived from Monte Carlo simulation (6,000 permutations)-based multiple comparison correction (Hagler, Saygin, & Sereno, 2006) using a cluster-forming threshold of  $p < .005$  and  $p < .001$  for thickness and surface area, respectively, in accordance with suggestions in Greve and Fischl (2018). Assuming that a cluster-wise corrected  $p$ -value of  $< .05$  is desired, the cluster size limits are 306 and 212 mm<sup>2</sup> for a cluster forming threshold of  $p < .005$  for the left and right hemisphere, respectively. For a cluster-forming threshold of  $p < .001$ , the corresponding sizes are 160 and 137 mm<sup>2</sup>. Results were projected onto the pial surface of *FsAverage* for display purposes. All other statistical analyses were performed using R (R Core Team, 2018). Tables over cortical thickness and

surface area for language-related cortical areas and their correlation statistics (Pearson's correlation with eICV and age as covariates) are provided as Supporting Information. The significance threshold from diffusion parameter correlation tests was Bonferroni corrected for multiple comparisons across the four analyzed white matter tracts, yielding a significance threshold of  $p = .05/4 = 0.0125$ .

## 3 | RESULTS

### 3.1 | LLAMA score interdependence

LLAMA score interdependence and dependence on working memory, fluid intelligence, musicality, and number of learnt languages is presented in Table 1, Spearman's  $\rho$  and uncorrected  $p$ -values. Results showed a weak correlation between LLAMA F and fluid intelligence (Raven score) and a stronger correlation between RAVEN score and number of learned languages. Taken together, the results provided evidence that differences in LLAMA tests within the group of participants did not reflect differences in other cognitive capacities and that LLAMA scores did not correlate with each other, that is, that they captured separate components of language-learning aptitude. Boxplots of the LLAMA scores are included in Supporting Information.

### 3.2 | White matter tract segmentations

Mean values with standard deviations for all diffusion parameters are presented in Table 2. MD, RD, FA, MK, and RK have been shown to vary between 0.80–0.93  $\mu\text{m}^2/\text{ms}$ , 0.31–0.70  $\mu\text{m}^2/\text{ms}$ , 0.41–0.83, 0.81–1.32, and 1.02–2.54, respectively, in healthy human white matter of the brain (Lätt et al., 2013). While parameter values were comparable to previous estimates in healthy human brains in AF and SLF

**TABLE 1** Spearman's rho and uncorrected  $p$ -values for comparisons between each background measure

	LLAMA B	LLAMA D	LLAMA F	OSPAN	Raven	Musicality general sophistication	Number of learned languages
LLAMA B (vocabulary)	1	$\rho = -0.15$ $p = .27$	$\rho = -0.11$ $p = .40$	$\rho = 0.25$ $p = .062$	$\rho = 0.01$ $p = .94$	$\rho = 0.12$ $p = .38$	$\rho = -0.13$ $p = .34$
LLAMA D (phonetics)		1	$\rho = -0.09$ $p = .51$	$\rho = 0.09$ $p = .51$	$\rho = 0.12$ $p = .38$	$\rho = -0.08$ $p = .54$	$\rho = 0.12$ $p = .38$
LLAMA F (grammar)			1	$\rho = -0.11$ $p = .41$	$\rho = 0.29$ $p = .032^*$	$\rho = 0.04$ $p = .78$	$\rho = 0.19$ $p = .15$
OSPAN				1	$\rho = 0.17$ $p = .22$	$\rho = 0.19$ $p = .16$	$\rho = 0.28$ $p = .034^*$
Raven					1	$\rho = 0.00$ $p = .98$	$\rho = 0.36$ $p = .0063^{**}$
Musicality general sophistication						1	$\rho = 0.06$ $p = .63$
Number of learned languages							1

Note: Dark cells are nonsignificant.

\* $p < .05$ .

\*\* $p < .01$ .

**TABLE 2** Mean diffusion parameter values for each studied white matter tract

Diffusion parameter	MD	FA	AD	RD	MK	AK	RK
Left AF	0.89 ± 0.03	0.36 ± 0.02	1.23 ± 0.04	0.72 ± 0.04	1.14 ± 0.04	0.93 ± 0.03	1.41 ± 0.08
Right AF	0.88 ± 0.03	0.36 ± 0.03	1.21 ± 0.03	0.71 ± 0.04	1.12 ± 0.05	0.95 ± 0.04	1.33 ± 0.10
Left SLF III	0.88 ± 0.03	0.38 ± 0.02	1.23 ± 0.03	0.71 ± 0.04	1.13 ± 0.05	0.91 ± 0.04	1.38 ± 0.08
Right SLF III	0.89 ± 0.03	0.36 ± 0.02	1.22 ± 0.03	0.72 ± 0.03	1.12 ± 0.05	0.94 ± 0.04	1.35 ± 0.11
Left UF	1.26 ± 0.24	0.22 ± 0.04	1.57 ± 0.29	1.10 ± 0.20	0.71 ± 0.11	0.76 ± 0.12	0.66 ± 0.11
Right UF	1.11 ± 0.12	0.23 ± 0.03	1.38 ± 0.14	0.97 ± 0.09	0.66 ± 0.06	0.70 ± 0.06	0.63 ± 0.10

Note: Mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD) are given in  $\mu\text{m}^2/\text{ms}$  while fractional anisotropy (FA), mean kurtosis (MK), axial kurtosis (AK), and radial kurtosis (RK) are dimensionless.

**TABLE 3** Statistics for cluster in left posterior inferior precuneus in which cortical surface area correlated with LLAMA B score

Cortical areas	Cluster size ( $\text{mm}^2$ )	Pearson's $r$	Peak Z	Peak Z coordinates (MNI $x, y, z$ )	Cluster-wise $p$
L. Precuneus	433	.473	5.11	-7, -59, 13	0.00200

III, values for UF were unfeasible for white matter tracts. Hence, values from UF were excluded from further analyses.

### 3.3 | LLAMA B

Cortical surface area in the left posterior inferior precuneus correlated with LLAMA B scores, as shown in Table 3 and Figure 2. The cluster-wise  $p$  in Table 3 is the corrected  $p$ -value after correction for multiple comparisons and is specific to the cluster size. No correlations were found between any diffusion parameter in any tract and the LLAMA B score (see Supporting Information for more information).

### 3.4 | LLAMA D

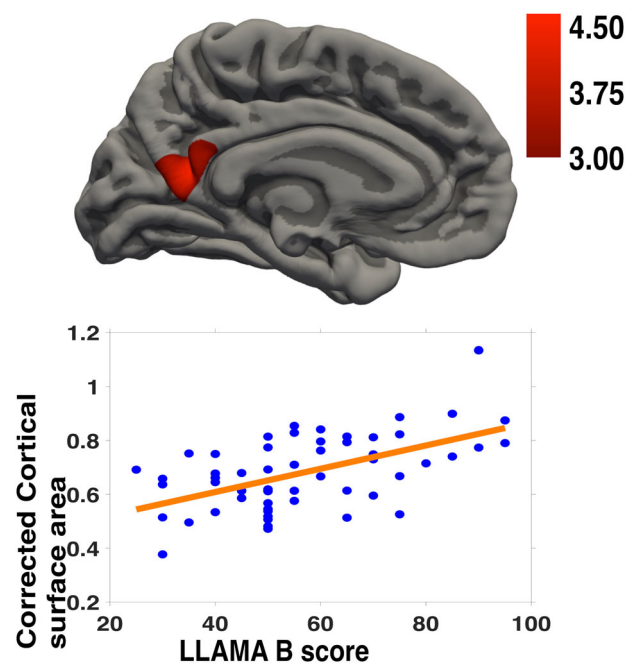
Mean AK along the left AF as well as mean MK, AK, and RK along the left SLF III correlated with phonetic memory (LLAMA D) score (Figure 3 and Table 4). This indicates that greater working memory capacity is associated with lower AK in the white matter tracts that are part of the dorsal language-processing stream. No correlations were found for the other tracts (see Supporting Information for details).

### 3.5 | LLAMA F

No correlations were found between cortical thickness, cortical surface area, or any diffusion parameter in any tract and LLAMA F score (see Supporting Information for more details).

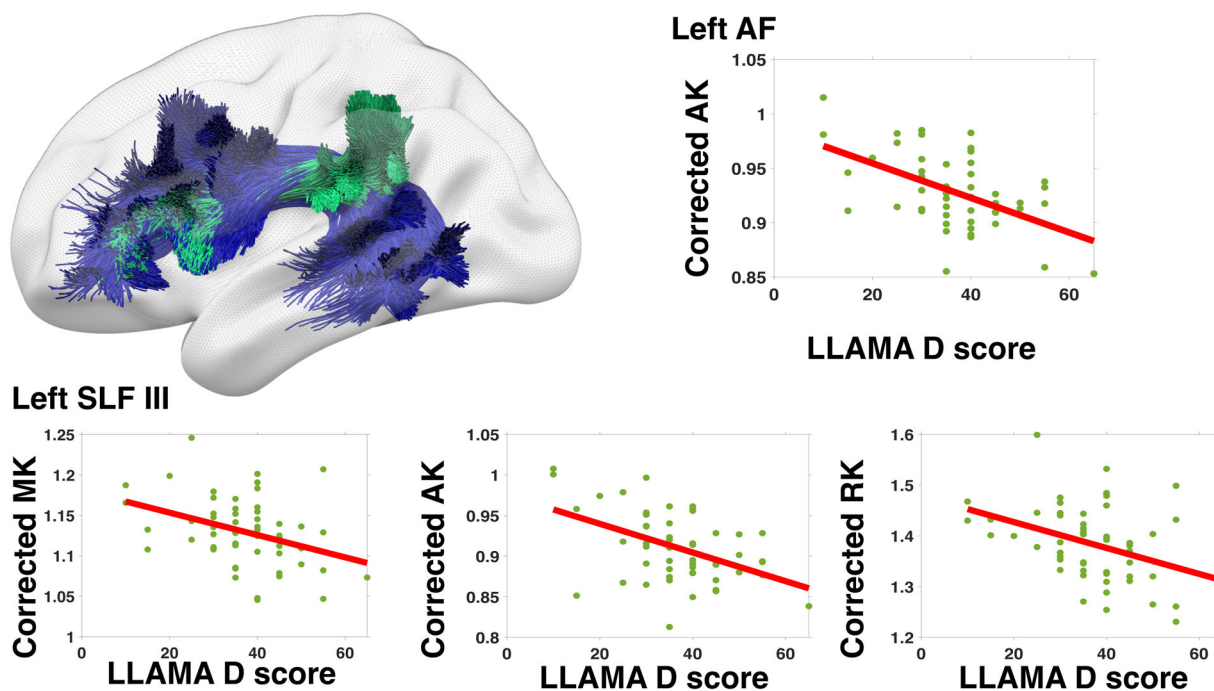
## 4 | DISCUSSION

This study investigated cortical morphometric and white matter fiber tract diffusion parameter correlations of language-learning aptitude. By finding neural foundations for language-learning abilities, we learn



**FIGURE 2** Vocabulary learning aptitude score (LLAMA B) correlated with cortical surface area in a left posterior medial cluster covering the left inferior posterior precuneus, given eICV and age as covariates

more about how the brain's anatomy can reflect differences in cognitive abilities. We found that greater vocabulary learning aptitude is associated with a larger cortical surface area in a left posterior medial cluster, possibly related to declarative memory capacity (Cavanna & Trimble, 2006). Moreover, phonetic memory negatively correlated with AK in left AF and SLF III, as well as MK and RK in the left SLF III. A negative correlation with MK, AK, and RK can be thought of as phonetic memory benefitting from a more coherent and homogeneous nerve fiber tract. Our results corroborate some and contradict other previous findings. We did not find any correlations between grey matter morphology in right HG (referred to as transverse temporal gyrus



**FIGURE 3** Phonetic memory aptitude (LLAMA D) correlated with axial kurtosis (AK) in left arcuate fasciculus (AF) and left superior longitudinal fasciculus part III. AF (blue) and SLF III (green) are shown in the image to the left, created using SurfIce (freely available at <https://github.com/neurolabusc/surf-ice>)

White matter tract	Diffusion parameter	Pearson's <i>r</i>	<i>t</i> (51)	<i>p</i>
L. AF	AK	-.518	-4.277	8.53e-05
L. SLF III	MK	-.356	-2.696	.0095
	AK	-.440	-3.462	.0011
	RK	-.378	-2.883	.0058

**TABLE 4** Correlation coefficients, *t*- and uncorrected *p*-values for significant correlations between diffusion kurtosis parameters and phonetic memory aptitude (LLAMA D)

Note: Bonferroni-corrected significance threshold is  $p < .0125$ .

in our Supporting Information) and LLAMA scores as expected from Turker et al. (2019). This could be due to the age differences between the groups investigated, as Turker et al. (2019) studied children between 10 and 16 years of age whereas our participants ranged between 20 and 27 years of age. This further suggests that the cortical drivers of language-learning aptitude are different in different stages of development. Also, while Turker and colleagues solely investigated HG, we performed a whole-brain analysis. If we assume a coarse parcellation of the cortex into 58 unique areas, used in one of the FreeSurfer parcellations that include HG (Desikan et al., 2006), then the Bonferroni corrected alpha-threshold for whole cortex analysis would be  $0.05/58 \approx 0.000862$ . For the correlations between grey matter volume in the right HG and LLAMA subtest scores, reported *p*-values range from .003 to .028. It should be noted that we found a nonsignificant correlation between cortical thickness in the right HG and LLAMA F score ( $r = .35$ , uncorrected  $p = .012$ ). See Supporting information for more details. Thus, the correlational results from Turker et al. (2019) would not survive correction for comparison across the entire brain surface. Taken together, the results found in the present study support

the claim that individual differences in language-related performance relate to individual differences in brain structure (Golestani, 2012).

#### 4.1 | Vocabulary learning aptitude correlates with cortical surface area in left inferior posterior precuneus

The vocabulary learning aptitude LLAMA subtest (LLAMA B) requires the test-taker to memorise the association of written words with cartoon figures. We found a correlation between this vocabulary learning aptitude score and the cortical surface area in a cluster in the left posterior inferior medial cortex referred to as the inferior posterior part of the precuneus.

The precuneus is involved in the integration between visuospatial inputs and declarative memory (Cavanna & Trimble, 2006), an essential step for LLAMA B test performance. The posterior precuneus is activated more when objects are correctly identified and connected to the correct source memory in episodic memory tests (Lundstrom,



Ingvar, & Petersson, 2005). Lundstrom et al. (2005) used a learning task in which participants were presented with a written noun and either a corresponding image or not. Concerning our results, this could indicate that those with high LLAMA B scores use of an episodic component to assist vocabulary learning and that vocabulary learning aptitude in part reflects this cognitive strategy. Left precuneus has a different cortical developmental trajectory in bilingual children and young adults, linked to, in part, increased demands on vocabulary knowledge (Pliatsikas et al., 2020). More left precuneal activity is elicited by remembered than forgotten words 1 week after practicing by naïve learners of Swahili in a vocabulary learning experiment (van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013). In sum, our results indicate that the cortical surface area of the left inferior posterior precuneus is of importance as regards the aptitude for learning new vocabulary. Relatively larger cortical surface area could speculatively indicate a greater predisposition for storing word-figure associations.

We did not find support for a correlation between FA in SLF III, arguably the closest correspondence to the tract between BA47 and the parietal lobe reported in (Xiang et al., 2012). This could be explained by Xiang et al.' (2012) use of probabilistic tractography on seed regions derived from functional connectivity patterns of the components of Broca's area (Xiang, Fonteijn, Norris, & Hagoort, 2010). We instead chose to use an automatic segmentation tool to ensure the anatomical validity of the included tracts.

#### 4.2 | Phonetic memory benefits from a coherent and more homogeneous left AF and SLF III

Phonetic memory capacity (LLAMA D) score correlated negatively with AK in the left AF and SLF III as well as MK and RK the left SLF III. The AF connects the IFG with the middle frontal gyrus, the posterior superior temporal gyrus and the temporal occipital transition region, while the adjacent SLF III connects the supramarginal gyrus with the prefrontal and ventral premotor cortices. AK and RK are indexes of tissue complexity along and across the principal diffusion direction, that is, along the fibers in an ideal white-matter voxel, respectively. MK is the average apparent kurtosis along all diffusion encoding directions. Higher AK could be due to the presence of nonaxonal cell membranes, for example, glial cells, astrocytes, and oligodendrocytes (Hui, Cheung, Qi, & Wu, 2008) or tortuosity of the axons (Fieremans, Jensen, & Helpert, 2011). Higher RK could additionally mean differences in axonal radii (Fieremans et al., 2011). While it is important to note that the DKI parameters are nonspecific to microstructural features, an intuitive interpretation of the results would be that more coherent and more homogeneous left AF and SLF III are beneficial for phonetic memory. This would hold for the tissue complexity in the left SLF III overall, not just in the principal fiber direction. Both left AF and SLF III are part of the dorsal language processing stream and are involved in the repetition of especially pseudowords (Hickok & Poeppel, 2004; Saur et al., 2008).

#### 4.3 | Perceptiveness to and learning foreign speech sounds

The ability to repeat meaningless words could perhaps be correlated with the implicit memory for foreign speech sounds as both, to some extent, require perceiving and temporarily storing novel sounds. It has been shown that expert phoneticians have larger pars opercularis of the left IFG, connected to the AF, and higher probability of split as well as more white matter in HG, bilaterally (Golestani et al., 2011). Golestani et al. did not investigate the white-matter connections between these areas but the implicit connection between the temporal lobe and the inferior frontal cortex fits our results well, assuming that trained phoneticians have greater phonetic memory capacity. Moreover, learners who are faster at perceiving novel speech sound contrasts have more white matter in (especially left) parietal regions (Golestani, Paus, & Zatorre, 2002). Success in learning Mandarin Chinese words from listening to a short film has been seen to be associated with greater functional connectivity between the left supplementary motor area and precentral gyrus as well as the left insula and the left rolandic operculum (Veroude, Norris, Shumskaya, Gullberg, & Indefrey, 2010). The structural connection between the supplementary motor area and precentral gyrus is part of SLF III and our results correspond well with the suggested importance of the functional connection for phonological rehearsal, but not necessarily speech articulation. Taken together, our findings are well in accordance with previously reported associations between the morphology of cortical areas connected by the left AF and SLF III and perceptiveness and learning aptitude for foreign speech sounds.

#### 4.4 | Phonological working memory

Studies of neural correlates of phonological working memory have directly implicated cortical areas connected by the left AF and SLF III, fitting well with the results presented here. Active phonological working memory tasks activate left IFG and posterior superior temporal sulcus (STS) in a "maintenance" phase and bilateral intraparietal sulcus, IFG and STS during the "comparison and decision" phase (Strand, Forssberg, Klingberg, & Norrelgen, 2008). The task used in Strand et al. (2008) differed from LLAMA D in that subjects were asked to actively keep the stimulus syllable strings in memory. Phonological working memory load is correlated with recruitment of bilateral superior temporal gyrus, left planum temporale and left precentral gyrus (Scott & Perrachione, 2019). More specifically, the posterior-medial planum temporale has been suggested to be a critical region for phonological working memory, supposedly as a site for "phonological storage" (McGettigan et al., 2010). Patients suffering from conduction aphasia, strongly associated with damage of the left AF (Damasio & Damasio, 1980; Tanabe et al., 1987), typically exhibit impaired phonological memory (Bartha & Benke, 2003). Altogether, left AF and SLF III connect important cortical areas for phonological working memory and our results indicate the importance for coherency and

homogeneity of the white matter tracts for the ability to implicitly learn foreign speech sounds.

#### 4.5 | Lack of correlations with LLAMA F

We did not replicate the findings in (Novén et al., 2019) that LLAMA F score correlates with cortical thickness in the left IFG pars triangularis and rostral middle frontal gyrus. In fact, we found no correlations at all with cortical thickness in the present study; a table with mean values of cortical thickness and surface area in several language-relevant cortical areas are provided as Supporting Information. One important difference between this and our previous study is that the T1-weighted images are of higher resolution (0.8 mm isotropic versus 1 mm isotropic) and thereby have less partial volume effects. This study included a greater number of participants (54 as opposed to 32). While this increases statistical power and thus increases the likelihood of finding associations, it is possible that this sample of subjects included more participants prone to use a strategy that does not benefit from greater cortical thickness in the aforementioned areas. The ratio of participants that scored more than 75 on the LLAMA F subtest, a score considered unusually high (Meara, 2005), was greater in Novén et al. (2019) 46.8% compared to 31.5% in this study. However, the LLAMA F test scores did not differ significantly ( $t[62.8] = -0.932$ ,  $p = .355$ , Welch two-sample *t*-test) between this study ( $M = 63.5$ ,  $SD = 19.7$ ) and Novén et al. (2019) ( $M = 67.9$ ,  $SD = 22.8$ ). Future studies could include a post-test survey to identify if there are differences in strategies to solve the task between subjects. Xiang et al. (2012) found correlations between grammatical inferencing aptitude and left-lateralization of the connection between BA45 and the posterior temporal lobe as well as the sum of the number of streamlines between BA6 and the posterior temporal lobe from both hemispheres. We found no correlations between any diffusion parameters in AF, arguably the closest corresponding tract in this study, and grammatical inferencing aptitude scores. This might, again, be due to differences in white-matter segmentation strategies.

Grammatical inferencing aptitude correlated with fluid intelligence but not with working memory, musicality, or number of learned languages. This replicates a known association between grammatical inferencing and intelligence (Cox, Lynch, Mendes, & Zhai, 2019). No other language-learning aptitude score correlated with any background or cognitive measure. This provides evidence that the LLAMA test battery measures language-learning specific cognitive constructs, not dependent on other aspects of cognitive capacity except for a moderate link between fluid intelligence and grammatical inferencing score.

#### 4.6 | Limitations and future directions

We identified five primary types of limitations in the present study. First, we did the imaging at an ultra-high field strength of 7T. While

the ultra-high field generally generates higher SNR and CNR, the magnetic field is also less homogeneous and susceptibility artifacts are more pronounced (van der Kolk, Hendrikse, Zwanenburg, Visser, & Luijten, 2013). However, we made substantial efforts to mitigate the effects of field inhomogeneities to make the results more reliable. FreeSurfer performs well at higher resolution but thickness estimates tend to decrease except in the cingulate and calcarine sulci as well as in the posterior bank of the central sulcus (Zaretskaya et al., 2018). The mitigated partial volume effects should improve the cortical thickness and surface area estimations. Maps of the mean and standard deviations of cortical thickness estimates from the data in this study and Novén et al. (2019) are given as Supporting Information. The present results should thus be more reliable than results from data with lower resolution.

The second limitation concerns the white matter segmentation. The TractSeg tool we used for this purpose was trained on data from lower field-strength MRI. However, a visual inspection of the tract segmentations indicates good performance on the 7T data for AF and SLF III (Figure 3). Still, the segmentation of IFOF did not map well to the DKI volume and diffusion parameters from UF were unfeasible and thus excluded. This is probably due to imaging artifacts in the inferior part of the DKI images. There could, therefore, exist correlations between the tissue microstructure of these tracts and language-learning aptitude scores to be discovered in future studies using methods that mitigate these issues. The slightly lower mean FA values found for AF and SLF III in this study probably have to do with the TractSeg segmentations being more generous than the manually positioned ROIs in Lätt et al. (2013).

The third limitation is that the LLAMA tests' ability to capture neuro-relevant variance depends on how well they match the granularity of the role of the cortical area/fiber tract. Future work could fine-tune the language-learning aptitude tests to divide the tests into components that match the functional roles of the different cortical areas. The validity of such tests must be motivated by functional studies (not necessarily fMRI). Nevertheless structural MRI could continue to capture brain morphological differences due to heightened acuity/apptitude for what the component captures.

The fourth limitation is that this study could potentially be limited in statistical power. This could possibly explain the few significant correlations found relative to the number of tests performed. A reasonable estimate (based on the studies cited in this article) of the correlation coefficients we can expect in correlations between behavioral measures and diffusion parameters or cortical morphometrical measures range from  $r = .35$ – $.56$ . Given a significance threshold of .05, we thus need between 23 and 62 participants to detect such correlation at a power of .8. This is assuming only one ROI but it is normal to include at least the contralateral homologue of the ROI in neuroimaging studies. If the significance threshold is Bonferroni corrected according to the diffusion analysis in our study with a reasonable number of ROIs (0.05/4), then the number of subjects needs to increase to between 31 and 86. For cortical morphometry, the measures range from  $r = .46$  to  $.66$  but have to be corrected for at least 58 ROIs (see reasoning above), requiring a number of subjects

between 31 and 73. As we included 57 subjects, we are confident that our results, although validation is necessary, can increase our understanding of the neural underpinnings of language-learning aptitude value to the field. However, future studies could benefit from including more subjects.

The fifth limitation is that we only studied young adult subjects and that the white matter maturation might not be complete. Studies have found both DTI parameters (Chen, Zhang, Yushkevich, Liu, & Beaulieu, 2016) and MK (Falangola et al., 2008) to be relatively stable in early adults. However, the MRI techniques we deploy might lack sensitivity regarding differences in white matter maturation.

## 5 | CONCLUSIONS

In conclusion, our results suggested a benefit from larger cortical surface in a left posterior medial cluster for vocabulary learning aptitude, possibly reflecting a greater declarative memory storing capacity for linguistically relevant word-figure associations. Moreover, the kurtosis parameters in known dorsal language processing stream tracts correlated negatively with phonetic memory. This indicates an impact from the coherency and homogeneity of white matter tracts connecting well-known cortical areas responsible for phonological storage on phonetic working memory. Our findings add to the knowledge of how cortical thickness, cortical surface area, and tissue microstructure of white matter fiber tracts correlate with talent for learning languages. In a wider perspective, this also means that differences in aptitudes and talents could, at least partly, be due to differences in cortical morphometry or white matter microstructure.

## ACKNOWLEDGMENTS

This work was supported by Knut and Alice Wallenberg Foundation (Grant No. 2014.0139 and 2018.0454), Marcus and Amalia Wallenberg Foundation (Grant No. 2018.0021), and the Swedish Research Council (Grant No. 2018.00632). Lund University Bioimaging Center (LBIC), Lund University, Sweden, is gratefully acknowledged for providing experimental resources. Open access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in "Language Learning Aptitude dataset" at <https://openneuro.org/datasets/ds003508>, doi: 10.18112/openneuro.ds003508.v1.0.0.

## ORCID

Mikael Novén  <https://orcid.org/0000-0003-0256-0522>

## REFERENCES

Abrahamsson, N., & Hylténstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second*

- Language Acquisition*, 30(4), 481–509. <https://doi.org/10.1017/S027226310808073X>
- Ades-Aron, B., Veraart, J., Kochunov, P., McGuire, S., Sherman, P., Kellner, E., ... Fieremans, E. (2018). Evaluation of the accuracy and precision of the diffusion parameter ESTimation with Gibbs and Noise removal pipeline. *NeuroImage*, 183, 532–543. <https://doi.org/10.1016/j.neuroimage.2018.07.066>
- Andersson, J. L., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, 20(2), 870–888. [https://doi.org/10.1016/s1053-8119\(03\)00336-7](https://doi.org/10.1016/s1053-8119(03)00336-7)
- Bartha, L., & Benke, T. (2003). Acute conduction aphasia: An analysis of 20 cases. *Brain and Language*, 85(1), 93–108. [https://doi.org/10.1016/S0093-934X\(02\)00502-3](https://doi.org/10.1016/S0093-934X(02)00502-3)
- Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87–136). Pittsburgh: University of Pittsburgh Press.
- Carroll, J. B., & Sapon, S. (1959). *Modern language aptitude test*. San Antonio: TX: Psychological Corporation.
- Catani, M., Jones, D. K., & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, 57(1), 8–16. <https://doi.org/10.1002/ana.20319>
- Catani, M., & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, 44(8), 1105–1132. <https://doi.org/10.1016/j.cortex.2008.05.004>
- Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, 129(3), 564–583. <https://doi.org/10.1093/brain/awl004>
- Chen, Z., Zhang, H., Yushkevich, P. A., Liu, M., & Beaulieu, C. (2016). Maturation along white matter tracts in human brain using a diffusion tensor surface model tract-specific analysis. *Frontiers in Neuroanatomy*, 10(9), 1–18. <https://doi.org/10.3389/fnana.2016.00009>
- Cox, J. G., Lynch, J. M., Mendes, N., & Zhai, C. (2019). On bilingual aptitude for learning new languages: The roles of linguistic and nonlinguistic individual differences. *Language Learning*, 69(2), 478–514. <https://doi.org/10.1111/lang.12341>
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, 9(2), 179–194.
- Damasio, H., & Damasio, A. R. (1980). The anatomical basis of conduction aphasia. *Brain*, 103(2), 337–350. <https://doi.org/10.1093/brain/103.2.337>
- Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3), 968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>
- Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1), 1–15. <https://doi.org/10.1016/j.neuroimage.2010.06.010>
- Duyn, J. H. (2012). The future of ultra-high field MRI and fMRI for study of the human brain. *NeuroImage*, 62(2), 1241–1248. <https://doi.org/10.1016/j.neuroimage.2011.10.065>
- Falangola, M. F., Jensen, J. H., Babb, J. S., Hu, C., Castellanos, F. X., Di Martino, A., ... Helpert, J. A. (2008). Age-related non-Gaussian diffusion patterns in the prefrontal brain. *Journal of Magnetic Resonance Imaging: JMIR*, 28(6), 1345–1350. <https://doi.org/10.1002/jmri.21604>
- Fieremans, E., Jensen, J. H., & Helpert, J. A. (2011). White matter characterization with diffusional kurtosis imaging. *NeuroImage*, 58(1), 177–188. <https://doi.org/10.1016/j.neuroimage.2011.06.006>
- Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9(2), 195–207. <https://doi.org/10.1006/nimg.1998.0396>

- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., ... Dale, A. M. (2004). Automatically Parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1), 11–22. <https://doi.org/10.1093/cercor/bhg087>
- Flöel, A., de Vries, M. H., Scholz, J., Breitenstein, C., & Johansen-Berg, H. (2009). White matter integrity in the vicinity of Broca's area predicts grammar learning success. *NeuroImage*, 47(4), 1974–1981. <https://doi.org/10.1016/j.neuroimage.2009.05.046>
- Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology*, 23(2), 250–254. <https://doi.org/10.1016/j.conb.2012.10.002>
- Golestani, N. (2012). Brain structural correlates of individual differences at low-to high-levels of the language processing hierarchy: A review of new approaches to imaging research. *International Journal of Bilingualism*, 18(1), 6–34. <https://doi.org/10.1177/1367006912456585>
- Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron*, 35(5), 997–1010. [https://doi.org/10.1016/S0896-6273\(02\)00862-0](https://doi.org/10.1016/S0896-6273(02)00862-0)
- Golestani, N., Price, C. J., & Scott, S. K. (2011). Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *The Journal of Neuroscience*, 31(11), 4213–4220. <https://doi.org/10.1523/jneurosci.3891-10.2011>
- Granena, G. (2012). *Age differences and cognitive aptitudes for implicit and explicit learning in ultimate second language attainment*. (Doctoral), University of Maryland, Unpublished. Retrieved from <http://hdl.handle.net/1903/12743>
- Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA language aptitude tests *sensitive periods, language aptitude, and ultimate L2 attainment*: John Benjamins.
- Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(3), 311–343. <https://doi.org/10.1177/0267658312461497>
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), 63–72. <https://doi.org/10.1016/j.neuroimage.2009.06.060>
- Greve, D. N., & Fischl, B. (2018). False positive rates in surface-based anatomical analysis. *NeuroImage*, 171, 6–14. <https://doi.org/10.1016/j.neuroimage.2017.12.072>
- Hagler, D. J., Saygin, A. P., & Sereno, M. I. (2006). Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *NeuroImage*, 33(4), 1093–1103. <https://doi.org/10.1016/j.neuroimage.2006.07.036>
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1–2), 67–99. <https://doi.org/10.1016/j.cognition.2003.10.011>
- Hui, E. S., Cheung, M. M., Qi, L., & Wu, E. X. (2008). Towards better MR characterization of neural tissues using directional diffusion kurtosis analysis. *NeuroImage*, 42(1), 122–134. <https://doi.org/10.1016/j.neuroimage.2008.04.237>
- Hutter, J., Nilsson, M., Christiaens, D., Schneider, T., Price, A. N., Hajnal, J. V., & Szczepankiewicz, F. (2018). *Highly efficient diffusion MRI by Slice-interleaved Free-waveform Imaging (SIFI)*. Paper presented at the 26th Annual Meeting of ISMRM, Montreal, Canada.
- Im, K., Lee, J. M., Lyttelton, O., Kim, S. H., Evans, A. C., & Kim, S. I. (2008). Brain size and cortical structure in the adult human brain. *Cerebral Cortex*, 18(9), 2181–2191. <https://doi.org/10.1093/cercor/bhm244>
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), 825–841. [https://doi.org/10.1016/S1053-8119\(02\)91132-8](https://doi.org/10.1016/S1053-8119(02)91132-8)
- Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2), 143–156. [https://doi.org/10.1016/S1361-8415\(01\)00036-6](https://doi.org/10.1016/S1361-8415(01)00036-6)
- Jensen, J. H., & Helpert, J. A. (2010). MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR in Biomedicine*, 23(7), 698–710. <https://doi.org/10.1002/nbm.1518>
- Lätt, J., Nilsson, M., Wirestam, R., Ståhlberg, F., Karlsson, N., Johansson, M., ... van Westen, D. (2013). Regional values of diffusional kurtosis estimates in the healthy brain. *Journal of Magnetic Resonance Imaging: JMRI*, 37(3), 610–618. <https://doi.org/10.1002/jmri.23857>
- Li, S. (2014). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, 36(3), 385–408. <https://doi.org/10.1093/applin/amu054>
- Lundstrom, B. N., Ingvar, M., & Petersson, K. M. (2005). The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. *NeuroImage*, 27(4), 824–834. <https://doi.org/10.1016/j.neuroimage.2005.05.008>
- Makris, N., Kennedy, D. N., McInerney, S., Sorensen, A. G., Wang, R., Caviness, V. S., Jr., & Pandya, D. N. (2004). Segmentation of subcomponents within the superior longitudinal fascicle in humans: A quantitative, in vivo, DT-MRI study. *Cerebral Cortex*, 15(6), 854–869. <https://doi.org/10.1093/cercor/bhh186>
- Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2020). White matter microstructure predicts foreign language learning in army interpreters. *Bilingualism: Language and Cognition*, 23(4), 763–771. <https://doi.org/10.1017/S1366728920000152>
- Mårtensson, J., & Lövdén, M. (2011). Do intensive studies of a foreign language improve associative memory performance? *Frontiers in Psychology*, 2(12), 1–7. <https://doi.org/10.3389/fpsyg.2011.00012>
- McGettigan, C., Warren, J. E., Eisner, F., Marshall, C. R., Shanmugalingam, P., & Scott, S. K. (2010). Neural correlates of sublexical processing in phonological working memory. *Journal of Cognitive Neuroscience*, 23(4), 961–977. <https://doi.org/10.1162/jocn.2010.21491>
- Meara, P. (2005). *LLAMA language aptitude tests: The manual*. University of Wales Swansea: Lognostics.
- Meyer, M., Liem, F., Hirsiger, S., Jäncke, L., & Hänggi, J. (2013). Cortical surface area and cortical thickness demonstrate differential structural asymmetry in auditory-related areas of the human cortex. *Cerebral Cortex*, 24(10), 2541–2552. <https://doi.org/10.1093/cercor/bht094>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS One*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Nilsson, M., Szczepankiewicz, F., van Westen, D., & Hansson, O. (2015). Extrapolation-based References improve motion and Eddy-current correction of high B-value DWI data: Application in Parkinson's disease dementia. *PLoS One*, 10(11), e0141825. <https://doi.org/10.1371/journal.pone.0141825>
- Novén, M., Olsson, H., Helms, G., Horne, M., Nilsson, M., & Roll, M. (2021). Language learning aptitude dataset. *OpenNeuro*. <https://doi.org/10.18112/openneuro.ds003508.v1.0.0>
- Novén, M., Schremm, A., Nilsson, M., Horne, M., & Roll, M. (2019). Cortical thickness of Broca's area and right homologue is related to grammar learning aptitude and pitch discrimination proficiency. *Brain and Language*, 188, 42–47. <https://doi.org/10.1016/j.bandl.2018.12.002>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Olsson, H., & Helms, G. (2021). Bias field correction of MPRAGE by an external reference—The poor man's MP2RAGE. *arXiv:2104.14020 [physics.med-ph]*
- Panizzon, M. S., Fennema-Notestine, C., Eyer, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., ... Kremen, W. S. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex*, 19(11), 2728–2735. <https://doi.org/10.1093/cercor/bhp026>



- Pierpaoli, C., Jezzard, P., Basser, P. J., Barnett, A., & Di Chiro, G. (1996). Diffusion tensor MR imaging of the human brain. *Radiology*, 201(3), 637–648. <https://doi.org/10.1148/radiology.201.3.8939209>
- Pliatsikas, C., Meteyard, L., Verissimo, J., DeLuca, V., Shattuck, K., & Ullman, M. T. (2020). The effect of bilingualism on brain development from early childhood to young adulthood. *Brain Structure and Function*, 225(7), 2131–2152. <https://doi.org/10.1007/s00429-020-02115-5>
- Qi, Z., Han, M., Garel, K., San Chen, E., & Gabrieli, J. D. E. (2015). White-matter structure in the right hemisphere predicts Mandarin Chinese learning success. *Journal of Neurolinguistics*, 33, 14–28. <https://doi.org/10.1016/j.jneuroling.2014.08.004>
- Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1), 1–48. <https://doi.org/10.1006/cogp.1999.0735>
- Rodriguez, S. M., Archila-Suerte, P., Vaughn, K. A., Chiarello, C., & Hernandez, A. E. (2018). Anterior insular thickness predicts speech sound learning ability in bilinguals. *NeuroImage*, 165, 278–284. <https://doi.org/10.1016/j.neuroimage.2017.10.038>
- Rogers, V. E., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R. (2016). Testing aptitude: Investigating Meara's (2005) LLAMA tests. *EUROSLA Yearbook*, 16(1), 179–210. <https://doi.org/10.1075/eurosla.16.07rog>
- Rogers, V. E., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association*, 1(1), 49–60. <http://doi.org/10.22599/jesla.24>
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., ... Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), 18035–18040. <https://doi.org/10.1073/pnas.0805234105>
- Schilling, K. G., Daducci, A., Maier-Hein, K., Poupon, C., Houde, J. C., Nath, V., ... Descoteaux, M. (2019). Challenges in diffusion MRI tractography—Lessons learned from international benchmark competitions. *Magnetic Resonance Imaging*, 57, 194–209. <https://doi.org/10.1016/j.mri.2018.11.014>
- Scott, T. L., & Perrachione, T. K. (2019). Common cortical architectures for phonological working memory identified in individual brains. *NeuroImage*, 202, 116096. <https://doi.org/10.1016/j.neuroimage.2019.116096>
- Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3), 143–155. <https://doi.org/10.1002/hbm.10062>
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., ... Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(Suppl 1), S208–S219. <https://doi.org/10.1016/j.neuroimage.2004.07.051>
- Strand, F., Forssberg, H., Klingberg, T., & Norrelgen, F. (2008). Phonological working memory with auditory presentation of pseudo-words—An event related fMRI study. *Brain Research*, 1212, 48–54. <https://doi.org/10.1016/j.brainres.2008.02.097>
- Tanabe, H., Sawada, T., Inoue, N., Ogawa, M., Kuriyama, Y., & Shiraishi, J. (1987). Conduction aphasia and arcuate fasciculus. *Acta Neurologica Scandinavica*, 76(6), 422–427. <https://doi.org/10.1111/j.1600-0404.1987.tb03597.x>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Retrieved from <https://www.R-project.org/>
- Teeuwisse, W. M., Brink, W. M., & Webb, A. G. (2012). Quantitative assessment of the effects of high-permittivity pads in 7 Tesla MRI of the brain. *Magnetic Resonance in Medicine*, 67(5), 1285–1293. <https://doi.org/10.1002/mrm.23108>
- Tournier, J. D., Calamante, F., & Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, 35(4), 1459–1472. <https://doi.org/10.1016/j.neuroimage.2007.02.016>
- Turker, S., Reiterer, S. M., Schneider, P., & Seither-Preisler, A. (2019). Auditory cortex morphology predicts language learning potential in children and teenagers. *Frontiers in Neuroscience*, 13(824), 1–16. <https://doi.org/10.3389/fnins.2019.00824>
- Turker, S., Reiterer, S. M., Seither-Preisler, A., & Schneider, P. (2017). “When music speaks”: Auditory cortex morphology as a neuroanatomical marker of language aptitude and musicality. *Frontiers in Psychology*, 8(2096), 1–17. <https://doi.org/10.3389/fpsyg.2017.02096>
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6), 1310–1320. <https://doi.org/10.1109/TMI.2010.2046908>
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505. <https://doi.org/10.3758/BF03192720>
- Van de Moortele, P.-F., Auerbach, E. J., Olman, C., Yacoub, E., Ugurbil, K., & Moeller, S. (2009). T1 weighted brain images at 7 tesla unbiased for proton density, T2\* contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. *NeuroImage*, 46(2), 432–446. <https://doi.org/10.1016/j.neuroimage.2009.02.009>
- van den Broek, G. S. E., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *NeuroImage*, 78, 94–102. <https://doi.org/10.1016/j.neuroimage.2013.03.071>
- van der Kolk, A. G., Hendrikse, J., Zwanenburg, J. J., Visser, F., & Luijten, P. R. (2013). Clinical applications of 7 T MRI in the brain. *European Journal of Radiology*, 82(5), 708–718. <https://doi.org/10.1016/j.ejrad.2011.07.007>
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80, 62–79. <https://doi.org/10.1016/j.neuroimage.2013.05.041>
- Veroude, K., Norris, D. G., Shumskaya, E., Gullberg, M., & Indefrey, P. (2010). Functional connectivity between brain regions involved in learning words of a new language. *Brain and Language*, 113(1), 21–27. <https://doi.org/10.1016/j.bandl.2009.12.005>
- Vos, S. B., Tax, C. M. W., Luijten, P. R., Ourselin, S., Leemans, A., & Froeling, M. (2017). The importance of correcting for signal drift in diffusion MRI. *Magnetic Resonance in Medicine*, 77(1), 285–299. <https://doi.org/10.1002/mrm.26124>
- Vuoksimaa, E., Panizzon, M. S., Chen, C.-H., Fiecas, M., Eyer, L. T., Fennema-Notestine, C., ... Kremen, W. S. (2015). The genetic association between neocortical volume and general cognitive ability is driven by global surface area rather than thickness. *Cerebral Cortex*, 25(8), 2127–2137. <https://doi.org/10.1093/cercor/bhu018>
- Wasserthal, J., Neher, P. F., Hirjak, D., & Maier-Hein, K. H. (2019). Combined tract segmentation and orientation mapping for bundle-specific tractography. *Medical Image Analysis*, 58, 101559. <https://doi.org/10.1016/j.media.2019.101559>
- Wasserthal, J., Neher, P. F., & Maier-Hein, K. H. (2018a). *Tract Orientation Mapping for Bundle-Specific Tractography*. Paper presented at the Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Cham.
- Wasserthal, J., Neher, P. F., & Maier-Hein, K. H. (2018b). TractSeg—Fast and accurate white matter tract segmentation. *NeuroImage*, 183, 239–253. <https://doi.org/10.1016/j.neuroimage.2018.07.070>
- Wen, Z., Biedron, A., & Skehan, P. (2016). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, 50(1), 1–31. <https://doi.org/10.1017/S0261444816000276>
- Wong, F. C. K., Chandrasekaran, B., Garibaldi, K., & Wong, P. C. M. (2011). White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *The Journal of Neuroscience*, 31(24), 8780–8785. <https://doi.org/10.1523/jneurosci.0999-11.2011>
- Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadeh, A., Parrish, T. B., & Zatorre, R. J. (2008). Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral Cortex*, 18(4), 828–836. <https://doi.org/10.1093/cercor/bhm115>



- Xiang, H., Dediu, D., Roberts, L., Oort, E. v., Norris, D. G., & Hagoort, P. (2012). The structural connectivity underpinning language aptitude, working memory, and IQ in the Perisylvian language network. *Language Learning*, 62, 110–130. <https://doi.org/10.1111/j.1467-9922.2012.00708.x>
- Xiang, H., Fonteijn, H. M., Norris, D. G., & Hagoort, P. (2010). Topographical functional connectivity pattern in the perisylvian language networks. *Cerebral Cortex*, 20(3), 549–560. <https://doi.org/10.1093/cercor/bhp119>
- Zaretskaya, N., Fischl, B., Reuter, M., Renvall, V., & Polimeni, J. R. (2018). Advantages of cortical surface reconstruction using submillimeter 7T MEMPRAGE. *NeuroImage*, 165, 11–26. <https://doi.org/10.1016/j.neuroimage.2017.09.060>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Novén, M., Olsson, H., Helms, G., Horne, M., Nilsson, M., & Roll, M. (2021). Cortical and white matter correlates of language-learning aptitudes. *Human Brain Mapping*, 42(15), 5037–5050. <https://doi.org/10.1002/hbm.25598>