



LUND UNIVERSITY

Brain anatomical correlates of perceptual phonological proficiency and language learning aptitude

Novén, Mikael

2021

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Novén, M. (2021). *Brain anatomical correlates of perceptual phonological proficiency and language learning aptitude*. [Doctoral Thesis (compilation), General Linguistics]. Lund University.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

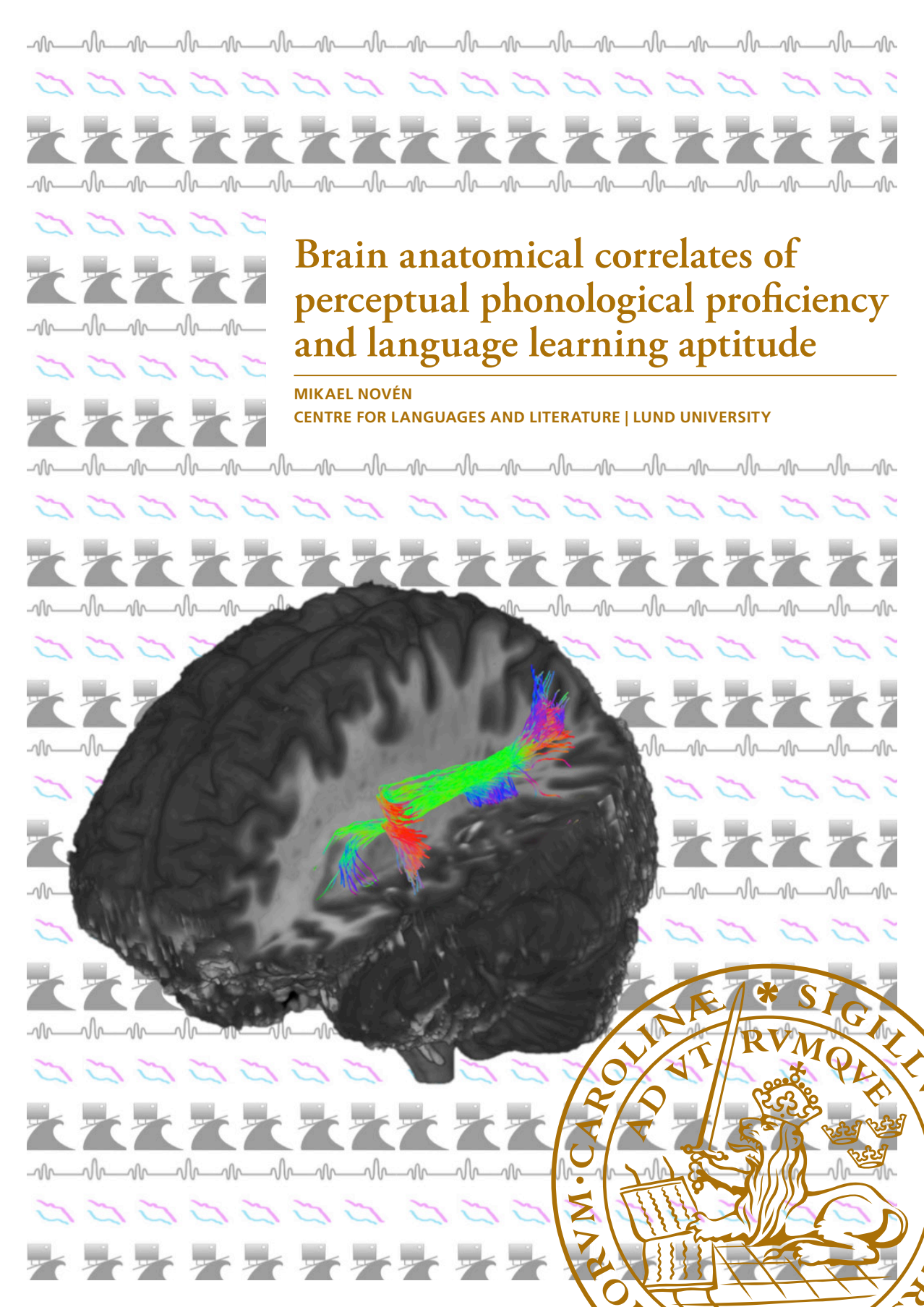
Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

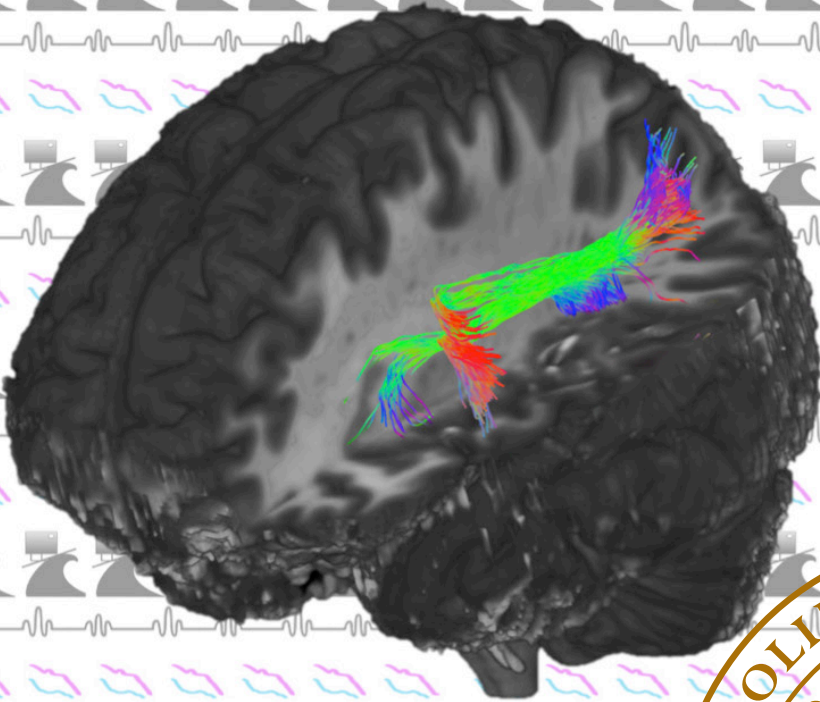
PO Box 117
221 00 Lund
+46 46-222 00 00



Brain anatomical correlates of perceptual phonological proficiency and language learning aptitude

MIKAEL NOVÉN

CENTRE FOR LANGUAGES AND LITERATURE | LUND UNIVERSITY





The Faculties of Humanities and Theology
Centre for Languages and Literature

ISBN 978-91-89213-43-2



9 789189 213432

Brain anatomical correlates of perceptual phonological proficiency
and language learning aptitude

Brain anatomical correlates of perceptual phonological proficiency and language learning aptitude

Mikael Novén



LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculties of Humanities and Theology,
Lund University, Sweden.

To be defended at LUX:B336. Date 2021-03-05 and time 10:00 (CET).

Faculty opponent

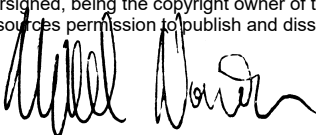
Professor Martin Lövdén

Department of Psychology, University of Gothenburg

Organization LUND UNIVERSITY Centre for Languages and Literature		Document name DOCTORAL DISSERTATION
Author Mikael Novén		Date of issue 2021-03-05
Title and subtitle Brain anatomical correlates of perceptual phonological proficiency and language learning aptitude		Sponsoring organization
Abstract The present dissertation concerns how brain tissue properties reflect proficiency in two aspects of language use: the ability to use tonal cues on word stems to predict how words will end and the aptitude for learning foreign languages. While it is known that people differ in their language abilities and that damage to brain tissue cause loss of cognitive functions, it is largely unknown if differences in language proficiencies correlate with differences in brain structure. The first two studies examine correlations between cortical morphometry, i.e. the thickness and surface area of the cortex, and the degree of dependency on word accents for processing upcoming suffixes in Swedish native speakers. Word accents in Swedish facilitate speech processing by having predictive associations to specific suffixes, (e.g. <i>fläck</i> _{accent1} + <i>en</i> 'spot+singular', <i>fläck</i> _{accent2} + <i>ar</i> 'spot+plural'). This use of word accents, as phonological cues to inflectional suffixes, is relatively unique among the world's languages. How much a speaker depends on word accents in speech processing can be measured as the difference in response time (RT) between valid and invalid word accent-suffix combinations when asked to identify the inflected form of a word. This can be thought of as a measure of perceptual phonological proficiency in native speakers. Perceptual phonological proficiency is otherwise very difficult to study, as most phonological contrasts are mandatory to properly interpret the meaning of utterances. Study I compares the cortical morphometrical correlates in the planum temporale and inferior frontal gyrus pars opercularis in relation to RT differences in tasks involving real words and pseudowords. We found that thickness of the left planum temporale correlates with perceptual phonological proficiency in lexical words but not pseudowords. This could implicate that word accents are part of full-form representations of familiar words. Moreover, for pseudowords but not lexical words, the thickness of the inferior frontal gyrus pars opercularis correlates with perceptual phonological proficiency. This association could reflect a greater importance for decompositional analysis in which word accents are part of a set of rules listeners need to rely on during processing of novel words. In study II, the investigation of the association between perceptual phonological proficiency in real words with cortical morphometry is expanded to the entire brain. Results show that cortical thickness and surface area of anterior temporal lobe areas, known constituents of a ventral sound-to-meaning language-processing stream is associated with greater perceptual phonological proficiency. This is consistent with a role for word accents in aiding putting together the meaning of or accessing a whole word representation of an inflected word form. Studies III and IV investigate the cortical morphometric associations with language learning aptitude. Findings in study III suggest that aptitude for grammatical inferencing, i.e. the ability to analytically discern the rules of a language, is associated with cortical thickness in the left inferior frontal gyrus pars triangularis. Furthermore, pitch discrimination proficiency, a skill related to language learning ability, correlates negatively with cortical thickness in the right homologue area. Moreover, study IV, using improved imaging techniques, reports on a correlation between vocabulary learning aptitude and cortical surface area in the left inferior precuneus as well as a negative correlation between diffusional axial kurtosis and phonetic memory in the left arcuate fasciculus and subsegment III of the superior longitudinal fasciculus. However, the finding correlation between cortical thickness and grammatical inferencing skill from study III was not replicated in study IV. Taken together, the present dissertation shows that differences in some language proficiencies are associated with regionally thicker or larger cortex and more coherent white matter tracts, the nature and spatial locus of which depend on the proficiency studied. The studies add to our understanding of how language proficiencies are represented in the brain's anatomy.		
Key words Prosody, word accent, MRI, language learning aptitude, cortical morphometry, diffusion kurtosis imaging		
Classification system and/or index terms (if any)		
Supplementary bibliographical information		Language English
ISSN		ISBN 978-91-89213-43-2
Recipient's notes	Number of pages 142	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature



Date 2021-01-25

Brain anatomical correlates of perceptual phonological proficiency and language learning aptitude

Mikael Novén



LUND
UNIVERSITY

Cover by Mikael Novén

Copyright Mikael Novén

Paper 1 © open access

Paper 2 © open access

Paper 3 © open access

Paper 4 © by the Authors (Manuscript unpublished)

The Faculties of Humanities and Theology
Centre for Languages and Literature

ISBN 978-91-89213-43-2 (print)

ISBN 978-91-89213-44-9 (PDF)

Printed in Sweden by Media-Tryck, Lund University, Lund 2021



Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN 

*To my father Lars who encouraged me to
use my head while helping me remember to
keep my feet rooted in the ground.*

Table of Contents

Acknowledgements	11
Populärvetenskaplig sammanfattning	13
Popular science summary	15
Abstract	19
List of original papers and contributions	21
Abbreviations	23
Introduction	25
Swedish word accents	26
Language learning aptitude	27
Background	29
Swedish word accents	29
Language Learning Aptitude	31
The LLAMA tests	32
Pitch Perception	33
Brain and language	34
Origins and measurements of neural activity	34
Cortical morphometry: thickness and surface area	35
Language processing streams	36
Magnetic Resonance Imaging	38
Nuclear magnetic resonance and imaging	38
MRI signals and tissue contrasts: Relaxation	38
Diffusion-weighted MRI and brain tissue microstructure	39
dMRI-based white matter tract segmentation	41
MR safety	41

Methods	43
Word accent processing task	43
LLAMA tests.....	44
Pitch Perception test	44
Participant background tests.....	45
Handedness	45
Musicality, working memory, and intelligence.....	45
Magnetic Resonance Imaging	46
Tissue segmentation and cortical morphometry.....	47
dMRI	50
White matter tract segmentation.....	50
Functional MRI	51
ERP.....	51
Power analysis.....	52
The investigations	53
Study I – Cortical thickness of planum temporale and pars opercularis in native language tone processing	53
Study II – Cortical thickness and surface area of left anterior temporal areas affects processing of phonological cues to morphosyntax.....	54
Study III – Cortical thickness of Broca’s area and right homologue is related to grammar learning aptitude and pitch discrimination proficiency.....	59
Study IV – Cortical and white matter correlates of language learning aptitudes	60
General discussion.....	61
Conclusions	65
Cortical morphometry in temporal speech processing areas is associated with Swedish phonological proficiency	65
Brain structure correlates with language learning aptitude	66
Outstanding issues and future directions	67
Word accents	67
Language learning aptitude	68
Are thicker cortices better for high-level cognition but thinner better for low-level cognition?	68
References	71

Acknowledgements

This thesis would never have been completed without the guidance and support from colleagues, family and friends. My main supervisor, Mikael Roll, has guided me through these years with great dedication, a seemingly endless supply of curiosity, and a healthy dose of humour. Our discussions have always made me wiser and more motivated to try to understand more about language and the brain. I owe much gratitude to my co-supervisors Merle Horne, Markus Nilsson, and Daniel Topgaard. Merle has helped me greatly to understand the field of linguistics and feel at home in research group and the department. Markus has been instrumental in my understanding of human magnetic resonance imaging and his razor-sharp intellect has really challenged me in ways that have made the articles he co-authored and myself as a researcher reach new potentials. Daniel has greatly helped me understand what life in academia is like. His insightfulness and infectious curiosity facilitated switching from studying napalm to study brains.

I have been truly lucky to get to work in a welcoming and bold group of amazingly talented researchers and great friends. A warm thank you to the neurolinguistic groups' past and present members: Andrea Schremm, Pelle Söderström, Sabine Gosselke Berthelsen, Frida Blomberg, Anna Hed, Otto Ewald, Andrea Fingerhut, Renata Kochančikaitė, and Anna Hjortdal. The discussions we have had at the department over the years have always been highly inspirational and helpful and the discussions in pub settings in Lund and all over the world even more so. Other colleagues at the department have made me feel very welcome and helped me understand the width of science, especially during intense lunch discussions. A special thank you goes out to the other doctoral students Sandra, Victor, Eline, Malin, and Niklas who have been very supportive and fun to work with.

A sincere thank you goes to Johan Mårtensson who gave extremely valuable feedback on my thesis as mock opponent and so much support and advice on life in and out of academia over the years. Magnus Lindgren gave me precious support when I needed it the most along with many insights in the mechanisms of the mind. Very often, I find myself grieving from the fact that I will not be able to work with him and that he will not be witness to the results of his encouragement: I made it through.

Lund university Bioimaging Centre, the National 7T Facility, and the MR-physics group are all gratefully acknowledged for all the help and support in setting up imaging sequences, experimental protocols, and participant logistics. It

has been a treat to be able to see the facility grow and to be a, hopefully not that annoying, test case for new routines. This dissertation would never be possible without the input of and support from Peter Mannfolk, Freddy Ståhlberg, Gunther Helms, Hampus Olsson, Björn Lampinen, Mads Andersen, Anders Sandberg, and Danielle Van Westen. A special thank you goes to Boel Hansson, Titti Owman, and Johanna Arborelius for your immense kindness and pragmatic problem solving skills.

Karin Markenroth Bloch has, apart from being an excellent site manager at the National 7T Facility, been a wonderful popular science presentation partner together with the brilliant Johan Mauritsson.

The wonderful people at RWI have really made the time of my doctoral studies more enjoyable. Thank you Kajsa, João, Greta, Samo, and Alexis for all wonderful insights and fun moments.

Thank you to friends and family who have made sure that life outside of office has been filled with joy, rest, and excitement. Thank you Fredrik for being there to support every step of this journey. Thank you Dennis, Björn, Filip, Jasper, and all other friends who have helped to distract me from worrying too much about work by inviting me to real and imagined adventures.

The love and help from my family is really what has held me up when things have been tough. My mother and brother have encouraged me at every step of the way and made sure that I have always had safe refuge to rest in throughout these years. Finally, Linnea, I love you with all I am and I know that we have each other's backs every minute of every day. Our sons, Olof and Albin, have heard very much about brain and language for all of their lifetime and made for excellent discussion partners during the writing of this dissertation. Olof, I am sorry it did not include more trains.

Populärvetenskaplig sammanfattning

Människans hjärna är ett otroligt komplicerat organ med ungefär hundra miljarder hjärnceller som skickar signaler mellan varandra i stora nätverk. På olika och ofta okända sätt gör dessa nätverk att vi kan föreställa oss saker, förstå vår omvärld och kommunicera med varandra. En sak som gör oss människor unika är vår välutvecklade språkförmåga. Vi kan berätta väldigt komplicerade historier för andra och lita på att de förstår vad vi menar eller kan fråga om detaljer. Dessutom kan vi lära oss helt nya språk även som vuxna. De här fantastiska färdigheterna gör kopplingen mellan hjärnan och språk intressant att studera. I den här avhandlingen visar jag några exempel på hur hjärnans struktur är förknippat med hur bra vi är på att använda och lära oss språk.

För att undersöka hur hjärnan ser ut har jag använt mig av en magnetkamera. I bilderna från magnetkameran syns olika vävnader i hjärnan väldigt tydligt. Det gör att vi lätt kan mäta tjockleken och ytarean på hjärnbarken, det yttre skiktet på hjärnan där hjärncellerna gör de flesta beräkningar och bearbetar information. Olika delar av hjärnbarken kommunicerar med varandra genom nervbanor. Nervbanor är samlingar av hjärncellernas långa utskott, kallade axoner, som kan leda signaler. En speciell magnetkamerateknik gör det möjligt att spåra hur vatten kan röra sig i hjärnvävnaden. Eftersom nervbanorna i princip består av rör som vattnet lätt kan röra sig längs med, men inte tvärs över, kan vi mäta två saker med hjälp av tekniken. För det första kan vi hitta och rita ut nervbanorna, något som inte är möjligt genom att bara titta på vanliga bilder. För det andra kan vi få en uppfattning om hur vävnaden ser ut på mikroskopisk nivå. Magnetkameran är alltså väl lämpad för att mäta hjärnans struktur.

Många språk använder ordmelodier för att göra skillnad på ords betydelse. På mandarin kan till exempel ordet ”*ma*” betyda ”mamma” eller ”häst” beroende på om det uttalas med en hög ton eller med en fallande och sen stigande ton. Andra språk kan göra skillnader på olika böjningsformer av ord med hjälp av toner. I svenskan har vi en liknande men ganska speciell användning av toner i ord. Tonerna kallar vi ordaccenter och ibland gör de skillnaden mellan betydelse av ord, till exempel ”*viner*” som i ”*vinden viner*” och ”*viner*” som i de röda dryckerna. Oftare använder vi dem dock för att förutsäga hur ord ska sluta. Om du lyssnar noga på när någon säger ”*bilen*” och ”*bilar*” kan du höra en skillnad i ordaccenter. Det är för att suffixet, ordändelsen, ”-en” kräver att ordet uttalas med en ordaccent medan suffixet ”-ar” kräver en annan. Exakt hur accenterna låter beror på vilken dialekt av svenska någon talar. Ordaccenterna är väldigt användbara eftersom de kan göra det lättare för oss att uppfatta vad folk säger.

Svenska modersmålstalare använder sig av ordaccenter olika mycket för att snabba på bearbetningen av talat språk och det kan vi mäta genom att se hur störda de blir om vi sätter ihop felaktiga kombinationer av ordaccenter och suffix. Med det måttet har jag undersökt om hjärnans struktur skiljer sig beroende på hur mycket vi använder oss av ordaccenter i språkbearbetningen. Jag upptäckte att det syns på hjärnbarken i vänstra tinningloben (precis ovanför örat, innanför tinningen) om någon förlitar sig mycket på ordaccenter för att förutsäga hur ett ord ska sluta. Hjärnbarken hos dem som förlitar sig mycket på ordaccenterna var både tjockare och hade större yta. Hjärnområdena i vänster tinninglob är förknippade med att tolka ljud och sätta ihop betydelsen av ord. Vi kan alltså se att mer hjärnbark i de här områdena är förknippat med att en person kan använda extra information i talet för att förstå vad som sägs bättre.

Förutom ordaccenter har jag också undersökt hur hjärnans struktur skiljer sig beroende på hur bra någon är på att lära sig nya språk. Tester som mäter denna förmåga är oftast byggda på fyra komponenter som John B. Carroll upptäckte på 1950-talet: 1) Hur bra en person är på att uppfatta språkljud och koppla ihop det med en betydelse. 2) Hur bra en person är på att memorera nya ord. 3) Hur bra en person är på att själv lista ut hur ett språk är uppbyggt utan att bli upplärd i språkets grammatik. 4) Hur bra en person är på att själv lista ut vilken grammatisk funktion ord och fraser har i ett språk. Jag mätte hur bra mina försöksdeltagare var på vardera komponent och undersökte hur det återspeglades i hjärnans struktur. Jag upptäckte tre sådana kopplingar. För det första såg jag att de som har lättare att lista ut hur språk är uppbyggda kan ha en tjockare hjärnbark i en del av Brocas område. Brocas område ligger lite framför och ovanför tinningloben och är förknippat med förmågan att sätta ihop meningar och att tala. Jag kunde dock inte få samma resultat i ett uppföljande experiment. Alltså behöver vi undersöka detta i fler studier för att förstå när hjärnbarkens tjocklek har betydelse för förmågan att lära sig nya språks uppbyggnad. För det andra upptäckte jag att personer som har lättare att lära sig nya glosor har hjärnbark med större yta i ett område som kallas vänster precuneus som har med minnet att göra. Precuneus ligger lite bakåt i mitten av hjärnan, nära nacken. Större yta i området kanske gör personer mer effektiva i att koppla ihop ord med betydelser. Till sist fann jag också att personer som har lättare att omedvetet lägga okända språkljud på minnet har mer ordnade nervbanor i vänster hjärnhalva. Detta kan betyda att personer som har bättre förbindelser mellan språkbearbetningsområden i hjärnan lättare intuitivt lägger nya språkljud på minnet.

I mina experiment behöver vi titta på många hjärnor för att se skillnader mellan personer som är olika bra på att använda eller lära sig språk. Vi kan alltså inte bara titta på en hjärna och säga om en person är bättre eller sämre på språk. Fynden jag presenterar i min avhandling hjälper dock till att bättre beskriva hur hjärnan gör det möjligt för oss att använda oss av och lära oss nya språk. Det kan i slutändan hjälpa oss att förstå mer av hur hjärnan fungerar, speciellt vad gäller språk som vi människor är alldeles särskilt duktiga på.

Popular science summary

The human brain is an incredibly complicated organ with about one hundred billion brain cells that send signals between each other in great networks. In various, and often unknown, ways, these networks make it possible for us to imagine things, understand our environment, and communicate with one another. One unique trait of humans is our well-developed language ability. We can tell each other complicated stories and trust that the other understands what we mean to say or can ask for details. Moreover, we can learn new languages, even as adults. These amazing skills make the connection between brain and language interesting to study. In this dissertation, I show some examples of how the brain's structure is associated with how good we are at using and learning language.

In order to investigate how the brain is structured, I have used a magnetic resonance imaging scanner. It is easy to distinguish between the different tissues of the brain in the images taken by the scanner. This makes it easy to measure the thickness and surface area of the cortex, the outer layer of the brain, where the brain cells make most of their calculations and information processing. Different parts of the cortex communicate via nerve tracts. Nerve tracts are bundles of the long outgrowths of brain cells, called axons, that can transport signals. A special magnetic resonance imaging technique allows us to track how water can move within the brain tissue. Because nerve tracts basically consist of tubes that water can easily move along, but not across, we can measure two things using the technique. First, we can find and segment the nerve tracts. This is not possible to do based on standard, anatomical, images. Second, we can get an idea of how the tissue is structured on the microscopic level. Accordingly, magnetic resonance imaging is well suited for measuring the brain's structure.

Many languages use pitch accents in order to discern the meanings of words. As an example, the Mandarin word "*ma*" can mean "mother" or "horse" depending on if it is pronounced with a high or a falling and then rising tone. Other languages mark different word inflections with tones. Swedish has a similar, but rather special, use of tones in words. The tones are called word accents and sometimes make the difference between different word meanings, for example the word "*viner*" can either mean "wines" as in "the wind wines" or as in the red beverages. More often, however, they are used to predict how a word will end. If you listen carefully as a Swedish native speaker says "*bilen*", the car, or "*bilar*", the cars, you can hear a difference in word accent. This is because the suffix "*-en*" demands that the word is pronounced with one word accent while "*-ar*" demands another. The

precise pronunciations of the word accents depend on which dialect of Swedish the speaker uses. Word accents are very useful as they can make it easier for us to comprehend what people are saying.

Swedish native speakers use word accents to varying degrees for speeding up speech processing and we can measure this by investigating how much bothered they are by invalid combinations of word accents and suffixes. Using this measure, I have investigated if the brain's structure differs depending on how much we use word accents in speech processing. I discovered that the dependency on word accents to predict how words will end is related to the structure of the cortex in the left temporal lobe (right above the ear, under the temple). The cortex in those who depend much on the word accents was both thicker and covered more surface area than those who did not depend on them as much. The brain areas in the left temporal lobe are associated with sound analyses and to construct the meaning of words. We can see that more cortex in these areas are associated with a person being able to use extra information in speech to better or more easily understand what is being said.

In addition to word accents, I have also studied how brain structure differs between people depending on how good they are at learning a new language. Tests that measure this ability is most often built on four components discovered by John B. Carroll in the 1950's: 1) How skilled a person is at perceiving a speech sound and associate it with a meaning. 2) How good a person is at memorizing new words. 3) A person's proficiency at uncovering and inferring the structure of a language without being explicitly taught in the grammar of the language. 4) How good a person is at figuring out the grammatical use for a word or phrase without explicit instruction. I measured how proficient my participants were at each component and investigated how this was reflected in the structure of the brain. I found three of those associations. First, I discovered that those who were better at inferring how languages are structured had a thicker cortex in a part of Broca's area. Broca's area lies a little bit in front of and above the temporal lobe and is associated with the ability to put together sentences and speaking. However, I could not find the same results in a subsequent study. This means that we need to further investigate when the thickness of the cortex is of importance for the ability to learn the structure of languages in more studies. Second, I found that people who are better at learning new vocabulary have a larger cortical surface area in an area called the left precuneus that has to do with memory. Precuneus is located a bit back and in the middle of the brain, close to the neck. Larger surface area in this area might make people more effective in associating words and their meanings. Lastly, I discovered that people who were good at unconsciously hearing and remembering foreign speech sounds had more ordered nerve tracts in the left hemisphere. This could mean that those who have better connections between speech processing areas of the brain more easily intuitively register new speech sounds.

In my experiments, we need to study many brains to see the differences between those who are good and those who are not so good at learning new languages. This means that we cannot just look at a brain and tell if a person is better or worse at using or learning language. However, the findings I present in my dissertation help to better describe how the brain makes it possible for us to use and learn languages. This could, in the end, aid our understanding of how the brain works, especially when it comes to language.

Abstract

The present dissertation concerns how brain tissue properties reflect proficiency in two aspects of language use: the ability to use tonal cues on word stems to predict how words will end and the aptitude for learning foreign languages. While it is known that people differ in their language abilities and that damage to brain tissue cause loss of cognitive functions, it is largely unknown if differences in language proficiencies correlate with differences in brain structure.

The first two studies examine correlations between cortical morphometry, i.e. the thickness and surface area of the cortex, and the degree of dependency on word accents for processing upcoming suffixes in Swedish native speakers. Word accents in Swedish facilitate speech processing by having predictive associations to specific suffixes, (e.g. *fläck_{accent1}+en* ‘spot+singular’, *fläck_{accent2}+ar* ‘spot+plural’). This use of word accents, as phonological cues to inflectional suffixes, is relatively unique among the world’s languages. How much a speaker depends on word accents in speech processing can be measured as the difference in response time (RT) between valid and invalid word accent-suffix combinations when asked to identify the inflected form of a word. This can be thought of as a measure of perceptual phonological proficiency in native speakers. Perceptual phonological proficiency is otherwise very difficult to study, as most phonological contrasts are mandatory to properly interpret the meaning of utterances. Study I compares the cortical morphometrical correlates in the planum temporale and inferior frontal gyrus pars opercularis in relation to RT differences in tasks involving real words and pseudowords. We found that thickness of the left planum temporale correlates with perceptual phonological proficiency in lexical words but not pseudowords. This could implicate that word accents are part of full-form representations of familiar words. Moreover, for pseudowords but not lexical words, the thickness of the inferior frontal gyrus pars opercularis correlates with perceptual phonological proficiency. This association could reflect a greater importance for decompositional analysis in which word accents are part of a set of rules listeners need to rely on during processing of novel words. In study II, the investigation of the association between perceptual phonological proficiency in real words with cortical morphometry is expanded to the entire brain. Results show that cortical thickness and surface area of anterior temporal lobe areas, known constituents of a ventral sound-to-meaning language-processing stream is associated with greater perceptual phonological proficiency. This is consistent

with a role for word accents in aiding putting together the meaning of or accessing a whole word representation of an inflected word form.

Studies III and IV investigate the cortical morphometric associations with language learning aptitude. Findings in study III suggest that aptitude for grammatical inferencing, i.e. the ability to analytically discern the rules of a language, is associated with cortical thickness in the left inferior frontal gyrus pars triangularis. Furthermore, pitch discrimination proficiency, a skill related to language learning ability, correlates negatively with cortical thickness in the right homologue area. Moreover, study IV, using improved imaging techniques, reports on a correlation between vocabulary learning aptitude and cortical surface area in the left inferior precuneus as well as a negative correlation between diffusional axial kurtosis and phonetic memory in the left arcuate fasciculus and subsegment III of the superior longitudinal fasciculus. However, the finding correlation between cortical thickness and grammatical inferencing skill from study III was not replicated in study IV.

Taken together, the present dissertation shows that differences in some language proficiencies are associated with regionally thicker or larger cortex and more coherent white matter tracts, the nature and spatial locus of which depend on the proficiency studied. The studies add to our understanding of how language proficiencies are represented in the brain's anatomy.

List of original papers and contributions

Study I

Schremm, A., Novén, M., Horne, M., Söderström, P., van Westen, D., & Roll, M. (2018). Cortical thickness of planum temporale and pars opercularis in native language tone processing. *Brain and Language*, 176, 42–47.

I contributed to conceptualising the study and ran cortical analyses, and extracted all cortical thickness and surface area estimates. I co-authored the manuscript.

Study II

Novén, M., Schremm, A., Horne, M., & Roll, M. (2021). Cortical thickness and surface area of left anterior temporal areas affects processing of phonological cues to morphosyntax. *Brain Research*, 1750, 147150.

I conceptualized the study, acquired all MRI data, and performed all data and statistical analyses. I was the main author of the manuscript.

Study III

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.

I conceptualized the study, acquired all data, and performed all data and statistical analyses. I was the main author of the manuscript.

Study IV

Novén, M., Olsson, H., Helms, G., Horne, M., Nilsson, M., & Roll, M. (manuscript). Cortical and white matter correlates of language learning aptitudes.

I conceptualized the study, acquired all data, and performed all data and statistical analyses. I was the main author of the manuscript.

Abbreviations

ADC	Apparent Diffusion Coefficient
AF	Arcuate Fasciculus
AD	Axial Diffusivity
AK	Axial Kurtosis
ANOVA	Analysis of Variance
DKI	Diffusion Kurtosis Imaging
dMRI	Diffusion-weighted Magnetic Resonance Imaging
DTI	Diffusion Tensor Imaging
EEG	Electroencephalography
ERP	Event-Related Potentials
FA	Fractional Anisotropy
fMRI	Functional Magnetic Resonance Imaging
HG	Heschl's Gyrus
HRF	Haemodynamic Response Function
IFG	Inferior Frontal Gyrus
IFGpo	IFG pars opercularis
IFGpt	IFG pars triangularis
IFO	Inferior Frontal-Occipital Fasciculus
LAN	Left Anterior Negativity
MD	Mean Diffusivity
MEG	Magnetoencephalography
MK	Mean Kurtosis
MLAT	Modern Language Aptitude Test
MRI	Magnetic Resonance Imaging
PrAN	Pre-Activation Negativity
PT	Planum Temporale
RD	Radial Diffusivity
RK	Radial Kurtosis
ROI	Region of Interest
RT	Response Time
SLF III	Superior Longitudinal Fasciculus subsegment III
UF	Uncinate Fasciculus

Introduction

Language in all forms needs to be interpreted by our brains in order to be meaningful. Perhaps it is fairer to say that acoustic signals broadcasted from the mouth of a speaker, writings on a piece of paper, or rapid hand gestures become language in that they are interpreted in the brains of the producer and receiver. Human language is remarkably complex, capable of transferring large amounts of information both through what is actually said, which needs to be decoded phonologically, morphologically, semantically, and syntactically, and what is not, which needs to be inferred pragmatically. The present dissertation investigates how the structure of the cortex and white matter tracts is related to two aspects of language use:

1. perceptual phonological proficiency, measured as the reliance on Swedish word accents as phonological cues to word endings, and
2. aptitude for learning new languages.

While we know that damage to the brain causes loss of function, much less is known about if and in what way differences in cognitive skills are manifested as differences in cortical morphometry, i.e. the thickness and surface area of the cortex, or as differences in tissue microstructure in white matter tracts, connecting functionally specified cortical areas. Most of what is known about how language is represented in the brain is learnt from either functional neuroimaging experiments or observations of how local brain damage (e.g. from stroke or neurodegenerative disease) impedes language use. Functional neuroimaging experiments can show what parts of the brain are more active when processing one stimulus than another or compared to rest. This means that measures of activity are not strictly quantitative; they are always measured relative to something. On the other hand, mapping brain areas to function based on observations on brain-damaged patients is a powerful study strategy. If loss of tissue equals loss on function, it provides evidence for a causal role for the damaged area in the studied behaviour, with the important limiting factor that the relevant behaviour has rarely been measured in subjects before they were damaged. Importantly, data from primary progressive aphasia (PPA), a neurodegenerative disease in which language capabilities become slowly but progressively impaired due to regional brain tissue loss, has greatly informed language processing theories (Gorno-Tempini et al., 2011). However, if atrophy, loss of tissue, means impaired function, perhaps, to some extent, the

reverse is true? Perhaps more tissue means better function, depending on the tissue and the function. In fact, general intelligence is known to correlate with regional grey matter volume or cortical thickness (Haier et al., 2004; Karama et al., 2009; Menary et al., 2013) as well as white matter tissue microstructure as measured by diffusion-weighted magnetic resonance imaging (dMRI; Malpas et al., 2016). Moreover, learning new languages is associated with local increases in cortical thickness (Mårtensson et al., 2012; Klein et al., 2014). If intelligence is reflected in cortical morphometry and white matter microstructure, it is possible that other cognitive skills are as well. With this as inspiration, we set out to investigate the cortical morphometric and white matter tissue microstructure correlates of perceptual phonological proficiency and language learning aptitude.

Swedish word accents

Swedish offers a special opportunity to study proficiency in perceiving and utilising phonological cues in speech processing. In almost all dialects of Swedish and Norwegian, words carry one of two possible word accents associated with the stressed syllable of a word (Bruce, 1977; Riad, 2014). An important and somewhat unique feature of word accents is that the choice of which word accent a word stem carries depends on how the word ends. Native speakers of Swedish use this fact to predict upcoming word endings, thus facilitating speech processing (Söderström et al., 2012; Roll et al., 2013; Roll et al., 2015; Söderström et al., 2016; Söderström et al., 2017b; Söderström et al., 2017a; Söderström et al., 2018). However, the word accents are almost never necessary for actually comprehending the meaning of words. This is an important difference to lexical tones in e.g. Mandarin or Thai where tonal features in words are mandatory for understanding what a word means. Thus, Swedish word accents can be used to measure degree of “perceptual phonological proficiency” in native speakers by measuring how impeded listeners’ processing become if the word accent and word endings are mismatched.

A measure of perceptual phonological proficiency could address some important limitations of previous studies on the neuroanatomical correlates of phonological processing, namely the focus on second language learning (Golestani et al., 2002; Golestani et al., 2007; Wong et al., 2011; Golestani, 2012; Sebastián-Gallés et al., 2012) or speech production (Schwartz et al., 2012). The processing investigated in those tasks is biased to involve the dorsal speech-processing route, responsible for mapping speech to motor representations (Hickok and Poeppel, 2004; Saur et al., 2008). In the case of second language learners, this route is prominent due to implicit repetition, requiring activation of motor representations of new speech sounds (López-Barroso and de Diego-Balaguer, 2017). This means that the involvement of the more meaning-oriented ventral stream (DeWitt and Rauschecker, 2012) is often obscured in studies of phonological processing. The

difference in dependence on word accents as phonological cues to suffix meanings might therefore add a more direct condition for studying the manifestation of perceptual phonological proficiency related to sound-meaning processing in the cerebral cortex. The hypothesis that the thickness and surface area of the cerebral cortex correlate with perceptual phonological proficiency in native speakers of Swedish is investigated in studies I and II.

Language learning aptitude

People vary in their ability to learn foreign languages. Researchers have attempted to measure talent for learning languages since at least the work by John B. Carroll in the 1950's leading to the design of the Modern Language Aptitude Test (MLAT; Carroll and Sapon, 1959). An important finding in the research behind the MLAT is that there are independent key components to language learning aptitude. In Carroll's framework these are phonetic coding ability, grammatical sensitivity, rote learning ability, and inductive language learning ability (Carroll, 1962). These components have been the basis for many adaptations of the MLAT, including the LLAMA test battery that is used in studies III and IV in this dissertation. The LLAMA test battery is free, computer-based and thus easy to administer (Meara, 2005). Analogous to intelligence mentioned earlier, language learning aptitude could perhaps manifest as differences in cortical morphometry or white matter tissue microstructure. The associations between brain structure and language-related performance, from low-level acoustic processing to high-level language control (Golestani, 2012), as well as how the brain changes from learning new language (Li et al., 2014) have been rather well studied. However, the investigations into brain structural correlates of language learning aptitude have been somewhat sprawling. Some have taken success in language courses as measures of learning aptitude. These have found that success in learning Mandarin in a classroom setting can be predicted by measures of white-matter microstructure (Qi et al., 2015) and that white-matter microstructure in frontal networks in conscript interpreters predicts how well they subsequently learn a foreign language (Mårtensson et al., 2020). However, general exam scores are very unspecific in terms of what language-related skills they measure and therefore the interpretation of the correlation with tissue microstructure is very diffuse and speculative. Another approach is to measure different rates of learning specific language features. For instance, some studies have focused on the predictors of learning to perceive or produce new speech sounds and found effects from the shape and size of Heschl's gyrus (HG; Wong et al., 2007a; Wong et al., 2008; Turker et al., 2017) and the cortical thickness of the anterior insula (Rodriguez et al., 2018). These results are interesting but limited to the investigation of one type of language learning aptitude. There are two studies that have investigated neuroanatomical correlates for the full range of language learning aptitudes, as measured by the

LLAMA test battery. One, looking only at white matter tissue microstructure in tracts derived from probabilistic tracking between manually defined seed points, ran the risk of investigating non-anatomical white matter tracts (Xiang et al., 2012). The other focused on the influence of the shape and volume of HG in children and adolescents (Turker et al., 2019). Both studies are described in more detail in the Background section.

In sum, there is a lack of studies that investigate the correlates of cortical morphometry on a range of language learning aptitudes, i.e. not limited to phonology, but more specific and related to language learning aptitude theory than general language course exam scores. Additionally, cortical surface area has been overlooked as a potential correlate of language learning aptitude. Moreover, correlates with white matter microstructure are based on simple imaging techniques giving crude characterisation of the tissues.

In study III, we therefore investigate possible cortical thickness correlates across the cortex with LLAMA subtest scores and in study IV we refine the design with higher-resolution anatomical data as well as tissue microstructure parameters derived from diffusion-weighted MRI. This allowed us to investigate not only the importance of the morphometry of the cortex but also the properties of the white matter fibre tracts connecting cortical areas. In study III, we also included a pitch perception test to bridge our understanding of perceptual phonological proficiency in Swedish native speakers with low-level pitch processing.

Background

Swedish word accents

In Swedish, each prosodic word carries one of two available word accents on the stressed syllable, named accent 1, also called “acute” accent, and 2, also called “grave” accent (Rischel, 1963; Bruce, 1977; Riad, 2014). The somewhat unique role of Swedish word accents in facilitating speech comprehension makes them an interesting feature to study to test theories of neural speech perception. In Central Swedish, the dialect of Swedish studied in this thesis, accent 1 is a low (L*) tone while accent 2 is a high (H*) tone (Bruce, 1987). Pitch accents mark prominence in all Germanic languages, but in Swedish and Norwegian, they also have lexical importance (i.e. they are related to meaning). Compared to lexical tones in e.g. Mandarin, word accents generally do not function to make categorical lexico-semantic distinctions, although some minimal pairs differing in only word accent do exist. They also do not mark grammatical distinctions as in, for instance, Somali (Le Gac, 2003). Instead they act as phonological cues to word endings (Rischel, 1963; Riad, 1998, 2012). As is presented in the pitch curve of the example sentence (shown in figure 1) *Kurt fick båt-en/ar till lunch* ‘Kurt got boat-the/s at lunch’, a monosyllabic word-stem such as *båt* ‘boat’ followed by a suffix marking singular definitive form *-en* carries accent 1, while if it is followed by, for instance, the plural suffix *-ar*, it carries accent 2. Correspondingly, present tense is associated with accent 1 and past tense with accent 2. However, accent 2 is associated with a wide range of word endings, on average 11 times more than accent 1 for a given word stem (Söderström et al., 2016). Native speakers of Swedish therefore use word accents to predict upcoming inflectional suffixes (Roll et al., 2010; Roll et al., 2013; Söderström et al., 2017a). Inflectional suffixes are morphemes added to words to change their grammatical form. Studies have consistently shown an advantage in response times for valid word accent-suffix associations in real nouns (Roll et al., 2013; Roll, 2015; Roll et al., 2015; Söderström et al., 2017a), pseudoword nouns (Söderström et al., 2017b), and verbs (Söderström et al., 2012).

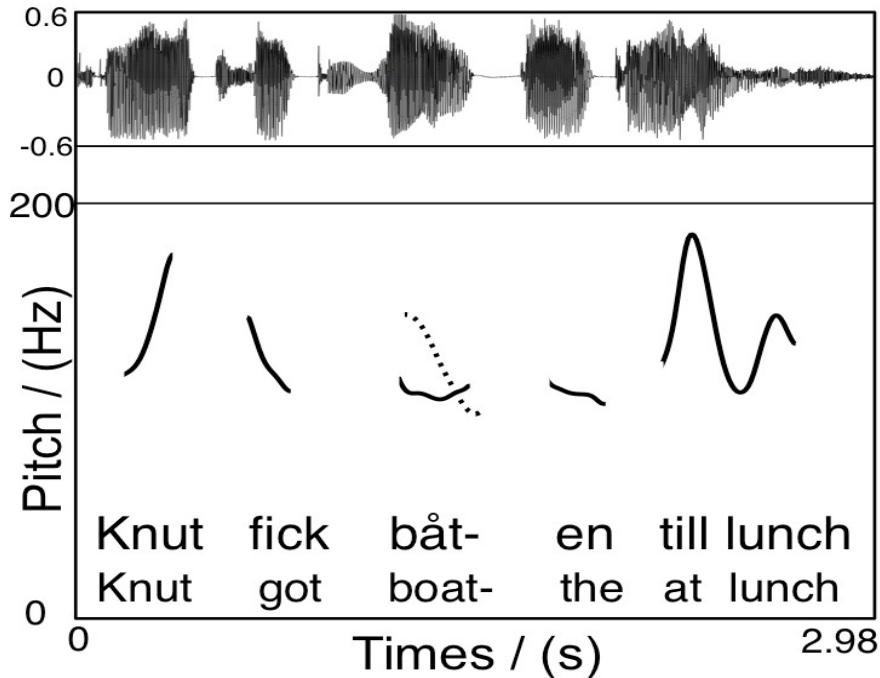


Figure 1: Soundwave (top) and pitch curve (bottom) of an example sentence with either accent 1 (low tone, solid line) or accent 2 (high tone, dotted line) on the target word *båt* (boat). Accent 1 is associated with the singular determiner suffix *-en* while accent 2 is associated with the plural suffix *-ar*.

The asymmetry in predictive power of accent 1 compared to accent 2 is indexed in the pre-activation negativity (PrAN) event-related potential (ERP) component (Roll, 2015; Roll et al., 2015; Söderström et al., 2016; Roll et al., 2017; Söderström et al., 2017b, 2018) as well as a greater P600 effect for accent 1 than accent 2 mismatches (Roll et al., 2010; Roll et al., 2013). The P600 is an event-related potential elicited by syntactic anomalies, possibly indicating a revision or reanalysis process (Osterhout and Holcomb, 1992). It could potentially be the case that the pitch height was somehow the cause behind the asymmetry in predictive value between the word accents. Importantly, however, this is not the case. The greater predictive value of accent 1 is present even in the South Swedish dialect where accent 1 is realised as a high tone and accent 2 as a low tone (Roll, 2015). Functional MRI (fMRI) experiments have shown that accent 1 yields more activity in a left temporal and inferiorfrontal cluster, including the left inferior frontal gyrus pars triangularis and opercularis as well as HG and superior and middle temporal gyri (Roll et al., 2015). The PrAN correlates with BOLD contrast in left HG, superior temporal, angular, and inferior frontal gyri, possibly indicating these areas as sources of the PrAN (Roll et al., 2015; Roll et al., 2017).

Invalid word accent-suffix combinations yield P600 effects (Roll et al., 2010; Roll, 2015; Söderström et al., 2017b; Söderström et al., 2017a) that correlate with increased BOLD activity in the left inferior parietal lobe (Roll et al., 2015). Bilateral supplementary motor areas and right middle frontal gyrus also yielded higher BOLD contrast for invalid than valid word accent-suffix combinations but did not correlate with the P600 effect. Investigations involving pseudowords have either found a left anterior negativity (LAN; Söderström et al., 2017a), reflecting increased morphological processing (Neville et al., 1991; Friederici et al., 1993), or an N400 (Söderström et al., 2017b), reflecting failed prediction (Kutas and Hillyard, 1980; Federmeier et al., 2007), elicited when processing invalid word accent-suffix combinations. Neither LAN nor N400 have been elicited in experiments on word accent processing in Central Swedish. Interestingly, L2 learners of Swedish trained specifically on using word accents as cues for suffixes show a LAN effect but no P600 (Hed et al., 2019), indicating that the learners registered the violation but were not proficient enough for the violation to cause a re-evaluation of the whole word.

The unique processing of word accents as phonological cues from the beginning of first language acquisition could have an impact on the cortical morphology in native speakers of Swedish. Moreover, because the accent information is not essential for understanding the meaning of words, there is a variance in the dependence on word accents as phonological cues in the population of native speakers. This dependence could be reflected as variations in the thickness and surface area of cortical areas relevant for speech processing. This hypothesis is investigated in studies I and II.

Language Learning Aptitude

As mentioned above, John B. Carroll developed the Modern Language Aptitude test (MLAT; Carroll and Sapon, 1959) in the 1950's partly as a way to identify gifted individuals the US army could educate in foreign languages for intelligence gathering purposes. He later synthesized an extremely influential model of language learning aptitude based on four factors: phonetic coding ability, rote learning ability, grammatical sensitivity, and inductive learning ability (Carroll, 1962). Phonetic coding ability means to be able to perceive a speech sound and associate a symbol to that sound. This is the basis for learning an orthographic system (i.e. learning to read and write). Rote learning ability essentially means the ability to learn and remember new words. Grammatical sensitivity and inductive learning ability refer to the abilities to infer the function of a lexical element and the rules governing the structure of a language, respectively, without specific knowledge of the language's grammar. Carroll's work has inspired a range of adaptations of the MLAT. Some have focused on different target groups while

others have added various cognitive components based on the same underlying cognitive factors (Wen et al., 2016).

Li (2014) did a meta-analysis of empirical research on language learning aptitude (measured as MLAT score) and its effect on second language (L2) grammar learning. He found an overall moderate association ($r=0.31$) between the two and that high school students were more likely to draw on aptitude than university students (Li, 2014). In the article, this is interpreted as the MLAT measuring aptitude for earlier, rather than later, stages of second language acquisition. However, an important caveat to that interpretation is that university students are a more selected group than high school students with less variation in aptitude.

In order to assess possible associations between brain structure and language learning aptitude, a more accessible and modern language learning aptitude test based on the theory supporting MLAT was used: the LLAMA tests.

The LLAMA tests

Meara (2005) developed the LLAMA tests as a free, easy-to-administer and L1-neutral language learning aptitude test loosely based on MLAT (Meara et al., 2001; Meara, 2005). It consists of a vocabulary (LLAMA B), a sound-symbol correspondence (LLAMA E), and a grammatical inferencing (LLAMA F) learning aptitude test. Moreover, the LLAMA tests also include a phonetic memory test (LLAMA D) to measure aptitude for implicitly remembering speech sounds that is not part of the components in MLAT. The details of how the tests are designed are given under Methods. One important limitation of the L1-neutrality is that the test stimuli in LLAMA B and LLAMA F subtests, presented shortly, use the Roman alphabet. It would in principle be possible to perform the tests without knowing some implementation of the Roman alphabet, as the sound-symbol correspondences are never needed for completing the tasks. However, it would surely be a different (less linguistic) test if the test-taker did not map the written word forms to sounds internally.

Two studies have investigated the neuroanatomical correlates of LLAMA test scores, specifically. One study focused on the shape of HG, also known as the transverse gyrus in the superior temporal plane, that commonly hosts the primary auditory cortex. HG is known to be very variable in shape across individuals and can either be a single gyrus or split into multiple gyri (Benner et al., 2017). Participants who score highly on the LLAMA tests typically have greater grey matter volume in and a higher probability of split right but not left HG than those with a low aptitude score (Turker et al., 2019). The second study used diffusion-weighted MRI to measure the importance of lateralisation of white matter tracts for LLAMA test scores and could show that LLAMA F scores correlated with greater fractional anisotropy (a parameter describing tissue microstructure explained in Methods) in a left hemispheric tract connecting Brodmann area (BA)

45 with the posterior temporal lobe, than in its right homologue, as well as with the sum of the number of streamlines in the connection between BA6 and posterior temporal lobe (Xiang et al., 2012). They also found that LLAMA E scores correlate with fractional anisotropy in a tract connecting BA45 in each hemisphere and that LLAMA B score correlated with difference in fractional anisotropy between the left and right tract between BA 47 and the parietal lobe. An important limitation of Xiang et al. (2012) study is that it relies on probabilistic tractography between seed points obtained from fMRI experiments. As such, the anatomical validity of these apparent connections cannot be assured (see motivation in Methods). The tracts reported in Xiang et al. (2012) are not reported in tractography atlases of white-matter tracts (Wakana et al., 2004; Catani and Thiebaut de Schotten, 2008). This limitation motivated the use of an anatomy-informed tool for white matter tract segmentation in study IV in the present dissertation.

There is on-going validation work done on the LLAMA tests and the results thus far summarized here. The scores are unaffected by gender or age (above 12 y.o.) but experience from formal education is significantly positively correlated with LLAMA B, E, and F but not D scores (Rogers et al., 2016; Rogers et al., 2017). The tests have been shown to be internally consistent and stable over time (Granena, 2013). LLAMA scores are positively correlated with pronunciation, lexis, and collocation learning scores in L2 learners with age of onset of learning between 16-29 years but not earlier for L1 speakers of Chinese learning Spanish (Granena and Long, 2013). This indicates that aptitude might be more important for adult than younger L2 learners. While these are promising results, more validation needs to be done, perhaps especially for the claim that the LLAMA tests are L1-independent. But overall, the scores are reliable enough to be used for investigating their neuroanatomical correlates and are therefore used in studies III and IV.

Pitch Perception

The ability to perceive pitch variations is important for language processing because pitch can convey prosodic information relating not only to semantics and grammar, but also to the emotional state of the speaker. A majority of the world's languages use pitch patterns to signal word meaning (Wong et al., 2008) or grammatical distinctions (Rolle, 2018) as well as morphosyntactic information (see section "Swedish word accents" above). Moreover, pitch perception ability predicts proficient learning of Spanish pronunciation in English-native learners (Posedel et al., 2011) and of English pronunciation in Japanese-native learners (Slevc and Miyake, 2006). Proficient pitch perception in musicians has also been shown to entail better encoding of linguistic pitch (Wong et al., 2007b).

Pitch perception skill can be measured either in relation to a musical framework (e.g. ability to name tones or intervals based on a musical notation system) or characterisation of the physical properties of the stimulus (e.g. how tones relate in terms of frequency). In this dissertation, we chose to measure pitch perception based on ability to detect pitch change direction, i.e. if a tone is higher or lower in pitch than another. This choice is motivated by us not wanting the test scores to be conflated with degree of formal music education.

There are some studies on the neural substrates of pitch perception skill. The right inferior frontal cortex is recruited when detecting pitch direction in speech syllables (Zatorre et al., 1992). Subjects suffering from congenital amusia (Peretz et al., 2002), a condition associated with impaired pitch perception, have shown to have damage to the right HG (Johnsrude et al., 2000) as well as a relatively thicker cortex in right frontal and temporal areas (Hyde et al., 2006; Hyde et al., 2007). Interestingly, the opposite is true for expert musicians, who instead have a relatively thinner cortex in right frontal areas (Bermudez et al., 2009). In a linguistic context, learning lexical tones lateralises their processing from the right to the left hemisphere (Lee et al., 2017).

Studies I and II investigate the importance of cortical thickness on linguistic tone pattern (word accent) processing. In order to relate the findings of those studies to the associations between non-linguistic pitch discrimination ability and cortical morphometry, a pitch perception test was implemented and the test scores used to correlate with cortical thickness in study III.

Brain and language

This section contains a brief overview of the relevant measures of cortical structure, brain tissue microstructure, and neural activity as well as some of the most influential theories on how the human brain comprehends and produces language.

Origins and measurements of neural activity

Brain tissue is made up of neurons and glial cells. Put very simply, the neurons take input in the form of neurotransmitter substances that attach to receptors on the dendrites that either excite or inhibit the neuron to transmit a signal (an action potential) via its axon. The whole signalling process from firing an action potential to re-establishing the membrane potential (i.e. the difference in concentration of ions between intracellular and extracellular water) takes around 4 ms (Purves et al., 2004). At the end of the axon are one or many end terminals called synapses that connect to the dendrites of other neurons. When the action potential reaches

the synapse, neurotransmitters are released onto the other neuron and cause ion channels to open in the postsynaptic cell membrane.

The subsequent postsynaptic change in potential is what is detectable in electrophysiological recordings such as electroencephalography (EEG). Action potentials are not thought to be visible in EEG because they occur over very small time-scales and the measured effects from action potentials from many neurons depend on the timing of the firing. If action potentials are fired out of sync, the potentials do not add together in the recordings. Postsynaptic potentials, on the other hand, last over several tens to hundreds of milliseconds and can thus be detected using EEG, provided there is synchronous activity from many neurons and favourable shape of the cortex (Luck, 2012). The action potential travels along the axon at a speed of about 5-25 m/s depending on the diameter of the axon (which varies between roughly 1-10 μm). However, this speed can be greatly increased to about 80-120 m/s if the axon is insulated with myelin sheets. Therefore, axon myelination is an important part of brain maturation, having its most dramatic phase up until about 2 years after birth.

The glial cells assist neural signalling by upholding the structure of the brain tissue, removing and recycling neurotransmitters, generating myelin sheets and similar tasks. Because neurons cannot store oxygen or glucose, the brain needs to be well supplied with blood from an extremely intricate and extensive network of blood vessels. In order to conserve energy, blood supply to brain areas that are not active is restricted until the need to resupply with oxygen and glucose arise, causing a peak in blood volume around five seconds after stimulus onset. The relatively close association between neural signalling and subsequent increase in blood supply is exploited in fMRI experiments in order to localise neural activity.

Cortical morphometry: thickness and surface area

The parts of the brain studied in this dissertation are limited to the cerebral cortex and cerebral white matter. The white matter mainly contains myelinated axons, mostly organised into bundles called tracts, which connect cortical areas across the brain. The high fat content gives the white matter its colour. The cerebral cortex constitutes the outer layer of the brain and contains mostly neuronal cell bodies, dendrites, and synapses (i.e. grey matter). During evolution, the brain has grown and has needed to reshape in order to fit more cortex into a smaller volume. This means that the brain has folded into ridges called gyri (singular: gyrus) and grooves termed sulci (singular: sulcus). Because the pattern of gyrification is very similar between humans, the gyri and sulci are used as anatomical landmarks, useful for co-registering brain images to each other in 3D-space (Fischl et al., 1999). Another popular framework for navigating around the brain is Brodmann areas, named after Korbinian Brodmann who parcellated the cerebral cortex on the basis of differences in organisation of the cells (Brodmann, 1909). According to the radial unit hypothesis, the cortex is built up by functionally specialised

columns (Rakic, 1988). This means that the thickness of the cortex is determined by how many cells are in the columns while the number of columns determines the cortical surface area. In fact, the cortical thickness and surface area are two independent measures of cortical morphometry (Meyer et al., 2013) with different genetic influences governing their development (Panizzon et al., 2009). The increase in cognitive ability seems more driven by increase in cortical surface area than cortical thickness (Vuoksimaa et al., 2015). In study III (Novén et al., 2019), we discuss the possibility that a thicker cortex is beneficial for more complex cognitive tasks, while a thinner, more specialised cortex could be beneficial for simpler computational tasks. This is discussed in more detail in the conclusions.

Language processing streams

Insights into how language is represented in the brain have been gained from studies on aphasia (impairments in comprehension or production of language) as well as neuroimaging experiments. While the classical Broca-Wernicke-Lichtheim-Geschwind-model (see for instance (Graves, 1997)) is considered outdated due to its focus on language as a unique cognitive process and its failure to encompass many new anatomical structures and functional mechanisms relevant for language processing (Poeppel, 2014; Tremblay and Dick, 2016), it is important to acknowledge that it has had a major impact as a groundwork to understanding the causes of aphasia. The following section describes the most influential models of how the brain is assumed to process aspects of language relevant for the studies in this thesis.

Speech processing has been observed to split into two streams in the brain, one ventral and one dorsal (Hickok and Poeppel, 2004; DeWitt and Rauschecker, 2012, 2013; Friederici and Gierhan, 2013; DeWitt and Rauschecker, 2016). This dual stream framework is inspired by the finding that the visual sensory input is processed in two separate streams, one that interfaces with motor systems for visually guided behaviour (e.g. reaching and grasping) called the dorsal (“how” or “action”) stream and one that is associated with shape and object recognition called the ventral (“what”) stream (Goodale and Milner, 1992). While all dual stream models for auditory language processing agree that a ventral stream is responsible for mapping sound to meaning, the dorsal stream is either given a rather restricted role as a basis for speech production and analysis of acoustic features (Hickok and Poeppel, 2007) or is also given a role in analysing complex syntactic structures (Friederici and Gierhan, 2013). This can be thought of as a disagreement regarding what the role of the left inferior frontal gyrus, canonical Broca’s area, does. While Hickok and Poeppel (2007) argue that it is simply a part of an articulatory network, others have highlighted its importance for also decoding the meaning of combinations of words and morphemes, suggesting a gradient from more semantically oriented processing in the inferio-anterior parts (pars orbitalis and pars triangularis) to more “pure” combinatorial syntactic

processing in the more posterior pars opercularis (Hagoort, 2005; Goucha and Friederici, 2015).

The place of input from phonological cues such as Swedish word accents that aid in speech processing in the dual stream models is unclear. However, investigations into the correlations between reliance on word accents as phonological cues to word endings (perceptual phonological proficiency) and cortical morphometry could provide additional information as the cortex could be reshaped during development depending on how well the accents are perceived and included in the analysis of the speech input. Moreover, studies on the neural bases for linguistic features that are not present in the languages that are studied as the basis for the theories of speech perception (mostly English, German, French, and Spanish) help test the generalizability of the theories.

Another suggestion for distinct processing of aspects of language comes in the form of the declarative/procedural model (Ullman, 2001, 2004). According to this model, the storage of memory representations of word forms (in the mental lexicon) is associated with the declarative memory system while the mental grammar, i.e. the basis for rule-based combination of lexical items, is part of the procedural memory system. The mental lexicon is a term for the mental representation of a word including its meaning, pronunciation, morphological and syntactic properties, and other relevant aspects of a word's meaning and use. The mental grammar, on the other hand, is the set of rules that allows for putting lexical items together to phrases and sentences. The declarative/procedural model is based on the words and rules theory (Pinker, 1999) that claims that the design of the language faculty is based on the distinction between the lexicon and grammar. In the case of English irregular and regular past tense verb inflections, irregular forms are thus suggested to be stored as whole forms, part of the lexicon and therefore part of the declarative memory system. On the other hand, rule-based decomposition is triggered by the past tense suffix in regularly inflected verbs and processed by the grammar in the procedural memory system. Similarly, Marslen-Wilson and Tyler (2007) suggest a core decompositional network linking left inferior frontal cortex with superior and middle temporal cortex, connected by the arcuate fasciculus, for processing regularly inflected words, while irregular inflections are stored as whole forms in the mental lexicon. However, unlike Ullman (2004), Marslen-Wilson and Tyler (2007) do not restrict the processing of inflectional morphology to the declarative/procedural memory dichotomy. Furthermore, they do not assume, as Pinker (1999) does, that the nature of neural computations differs depending on whether grammatical morphemes are present or not.

It is possible that Swedish word accents are processed analogously to the English irregular and regular past tense verb inflections. Word accents in very frequent words could then be stored together with the word stem/suffix combination in the mental lexicon, and thereby aid in lexical access, while in less frequent words they could aid in decompositional processing by predicting the probable suffix to end the word.

Magnetic Resonance Imaging

Nuclear magnetic resonance and imaging

Magnetic resonance imaging (MRI) is an imaging technique that utilizes magnetic properties of certain atomic nuclei to generate image contrast. The by far most commonly used nuclei to study in MRI are the hydrogen nuclei (protons) within water molecules as they are most abundant in the tissue. Put simply, each proton seemingly spins around its own axis and this generates a small magnetic moment. Under normal circumstances, i.e. outside of strong magnetic fields, the magnetic moments are randomly aligned and therefore the net magnetisation is zero. Inside an MRI scanner, however, there is a strong magnetic field, B_0 , and when a participant is placed within this magnetic field, the magnetic moments align in its direction. It is, however, possible to perturb the aligned magnetic moments using an electromagnetic field if it is in resonance with the precession (i.e. the change in orientation of the rotational axis of the spinning proton) of the magnetic moment of the nuclei. Importantly, the resonance frequency determines which nuclei are perturbed and thus generate a signal. The resonance frequency is specific for a given type of nucleus (with spin) in a magnetic field of a specific magnitude. The measured signal in MRI is the return of the magnetic moments to their aligned state, measurable as induced currents in receive coils around the participant. By applying magnetic field gradients across the participant, the resonance frequency changes locally, and thus, signal is only produced from a small image volume (voxel).

MRI signals and tissue contrasts: Relaxation

The return of the magnetic moments to their aligned state is characterised by two relaxation mechanisms: T1 and T2 relaxation. They each generate unique contrasts in the images. T1-relaxation depends on how efficient the interactions between the nucleus and its immediate environment are. T2 relaxation is instead due to small microscopic magnetic fields from surrounding molecules. The microscopic magnetic fields will cause the spins to precess out of phase and signal is lost. However, the measured T2 will also depend on the quality of the applied magnetic field gradients and differences in how much the tissue is affected by magnetic fields (magnetic susceptibility). The effective T2 that is measured if these effects are not corrected for is called T2*. T2* can sometimes be useful as it differs between oxygenated and deoxygenated blood. This difference can be used in fMRI experiments as an indirect measure of brain activity.

The relaxation times are different in different tissues, allowing for segmentation of images into different tissue types. In T1-weighted images, tissues with short T1 (e.g. white matter) are bright because it is possible to capture more of the return to

the aligned state. In T2-weighted images, tissues with short T2 (e.g. white matter) are dark because more of the signal is lost to T2-relaxation. Images can also be proton-density (PD) weighted. In PD-weighted images, the number of water molecules determines the image brightness. The difference in image brightness between tissue types makes it possible to segment the grey matter from the white matter and cerebrospinal fluid and measure the cortical thickness and surface area across the cortex.

Diffusion-weighted MRI and brain tissue microstructure

Diffusion-weighted MRI (dMRI) is based on encoding the random motion of water molecules in a tissue using magnetic field gradients. If one encoding gradient is applied followed by another of equal size but with switched polarity, the signal should be unaffected (the effect of the gradients cancel each other out). However, because of diffusion of the water molecules, the same hydrogen nucleus will not be exposed to exactly the same magnetic field change and the measured signal will therefore decrease. In dMRI experiments, we typically only study the self-diffusion and assume that there are no diffusion-driving differences for water molecules in the tissue. This is an appropriate assumption except for the flow of blood. However, blood flow operates at completely different time-scales than those studied in dMRI experiments and is thus negligible. If there are no barriers or concentration gradients in the tissue, the displacement of water is described by a Gaussian probability density function.

However, in brain tissues there are plenty of barriers for the water molecules. In any measurement, the apparent diffusion coefficient (ADC) will be lower than what would be expected for free water. For ease of thought, we can assume three typical voxels: one containing only cerebrospinal fluid, one containing cell bodies and dendrites, and the last containing only axons bundled in only one direction. These can be said to be typical cerebrospinal fluid, grey matter, and white matter voxels, correspondingly. Pure cerebrospinal fluid voxels contain no barriers for water diffusion and the displacement can therefore truthfully be described as a Gaussian. In pure grey matter voxels, there are plenty of barriers for the water molecules but they are not ordered in any particular direction and therefore the displacement is still described as a Gaussian but the mean displacement is shorter. However, in the pure white matter voxel, the water can diffuse more in one direction, the one along the axons, than in the other directions. If diffusion is equal in all directions it is said to be isotropic while if diffusion is greater in one direction than the others it is said to be anisotropic. Displacement of water in anisotropic diffusion systems can be described by the diffusion tensor (Basser et al., 1994a, b). The diffusion tensor is a 3x3 displacement covariance matrix that describes the mean displacement in the three principal laboratory axes (x,y,z), in the diagonal elements.

In order to facilitate interpretation and comparison, the diffusion tensor is often described using a number of scalars attempting to capture the most important properties of the (anisotropic) diffusion in a voxel. These scalars can be obtained through the eigenvalues of the diffusion tensor (Pierpaoli et al., 1996). A useful convention is to order the eigenvalues in falling order, $\lambda_1 > \lambda_2 > \lambda_3$. The eigenvalues correspond to the apparent diffusivities along the principal axes of diffusion in the voxel, the directions of which are given by the eigenvectors. In our imagined pure white matter voxel where all axons are aligned perfectly in one direction, $\lambda_1 \gg \lambda_2 > \lambda_3$, i.e. the anisotropy is very large. Contrarily, in the perfect CSF and grey matter voxels, $\lambda_1 \approx \lambda_2 \approx \lambda_3$, i.e. the diffusion is isotropic. Because λ_1 corresponds to the diffusivity along the main axis of diffusion, it is sometimes called the axial diffusivity (*AD*) or D_{\parallel} , and the mean of λ_2 and λ_3 is correspondingly called the radial diffusivity (*RD*) or D_{\perp} .

The mean diffusivity (*MD*), sometimes called the apparent diffusion coefficient (*ADC*), can be thought of as a mean diffusivity independent of direction. MD is calculated as

$$MD = ADC = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}.$$

Fractional anisotropy (*FA*) is a quantification of the anisotropy, i.e. how much better the water molecules can move in one direction compared to the others (Pierpaoli et al., 1996). It is computed as

$$FA = \sqrt{\frac{3((\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2)}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}.$$

If all eigenvalues are equal, the diffusion profile is isotropic and *FA* is 0. If, on the other hand, λ_1 is infinitely larger than the other two, *FA* is 1. There is eagerness in the research community to relate diffusion parameters to structure in the white matter. This causes many researchers to equate for instance *FA* with “white matter integrity” or some similar interpretation. It is very important to remain humble to the fact that there are many possible explanations to a given value of *FA* (except 1 that is an infinitely thin tube of a perfectly non-permeable material). The naming of axial and radial diffusivity also showcases the temptation to think of the ideal white-matter voxel to be the normal situation to study. It is not. It does, however, give measures related to the tissue microstructure within a voxel, just not in a clear-cut way.

More detailed information about the tissue microstructure can be obtained by using diffusion kurtosis imaging (*DKI*; Jensen et al., 2005; Hui et al., 2008; Fieremans et al., 2011), instead of the standard diffusion tensor imaging (*DTI*) described above. If the diffusion encoding is strong enough or the diffusion time

short enough, the multiple barriers in forms of cell membranes and macromolecules will impact the signal in a diffusion-weighted MR image. This causes the probability function to become less Gaussian. The shape of a probability distribution can be measured as the (excess) “kurtosis”. In order to calculate the kurtosis, at least two non-zero encoding strengths and 15 encoding directions are needed (Jensen et al., 2005). From the kurtosis, three parameters are commonly derived: Axial and radial kurtoses (AK and RK) that describe the diffusion kurtoses along or perpendicular to the principal diffusion direction and mean kurtosis (MK), the mean kurtosis across all encoding directions. To some extent, one can say that the greater the tissue complexity, the greater the value of the kurtosis parameter.

dMRI-based white matter tract segmentation

Beyond providing measures of tissue microstructure, it is possible identify cerebral nerve fibre tracts from dMRI data. This is possible because the water molecules can move more freely along than across the axons that make up the nerve fibre. Thus, fibres can be tracked based on the principal diffusion direction, i.e. the direction in which the water molecules can move most freely, in each voxel. This technique is called tractography (Basser et al., 2000) and is an important tool for mapping the structural connectome (i.e. the wiring diagram) of the brain (Sporns, 2011). If a voxel is a perfect set of myelinated axons aligned in one direction, one can be sure that following the principal diffusion direction means following the white matter fibre bundle. However, this is never really the case for imaging voxels that are large enough to contain at least ten thousand axons. In practice, the voxels also contain glial cells and the axons bend and cross, and while there are ways to mitigate this fact by estimating fibre orientation distributions, i.e. finding more than one major diffusion direction, for each ROI and using more robust probabilistic methods (Tournier et al., 2007; Tournier et al., 2012), tractography results are still not 100% reliable. Consequently, in order to ensure anatomical validity of the white-matter tracts one wishes to study, it important to add constraints on the tracts, either by manual editing or by using software tools.

MR safety

MRI is a safe neuroimaging technique that does not include ionising radiation, as in e.g. computer tomography. Nor does it require any contrast agent to be injected into the participant, as in e.g. positron emission tomography. Most MRI scanners in a hospital setting have B_0 ranging from 1.5 T to 3 T, with the current record for human MRI at 10.5 T (Sadeghi-Tarakameh et al., 2020). As a comparison, a fridge magnet is about 5 mT and the earth’s magnetic field between 30 and 60 μ T. Exposure to magnetic field (gradients) in MRI scanners has not been found to be a

significant risk to human tissues (Hartwig et al., 2009). The main risks involved in an MRI experiment instead concern effects on magnetic and/or conducting implants in the body of the participant or on objects in the vicinity of the scanner as well as tissue heating due to the energy deposit from the perturbing electromagnetic field (Panych and Madore, 2018). Tissue heating is limited by restrictions on the imaging protocols in the scanner. Implants and objects are routinely controlled to assess the safety of them entering the scanner room.

Participants might suffer some transient discomfort from short-term peripheral nerve stimulation and loud acoustic noises from the switching magnetic field gradients (Delfino et al., 2019). Moreover, movement into or inside of the static magnetic field can cause nausea and dizziness due to induced currents in the vestibular (balance) system (Roberts et al., 2011). These effects are fleeting and possible to mitigate using hearing protection and by moving slowly in and out of the static magnetic field.

Taken together, MRI is a safe technique suitable for studying healthy volunteers with no known long-term adverse effects as long as no magnetic or electrically conducting materials are allowed into the scanner room in an uncontrolled way.

Methods

Word accent processing task

Behavioural experiments were designed in order to investigate the impact of word accents on speech processing in studies I (Schremm et al., 2018) and II (Novén et al., 2021). This impact was used as a measure of the participants' perceptual phonological proficiencies. In all experiments, the stimuli were sentences with the form *Britt fick* [target] *på stan* 'Britt got [target] in town' spoken by a male native speaker of Central Swedish, where the target was a monosyllabic noun stem carrying either word accent 1 or 2 combined with either a plural or singular suffix. The name of the subject and the location in the prepositional phrase varied between sentences but were always monosyllabic. This yielded four sentences for each noun stem, one for each combination of word accent and suffix. Study I (Schremm et al., 2018) used data from two previously published studies to investigate cortical morphometrical correlates to word accent cue dependence. One of the studies concerned phonotactically legal pseudowords (Söderström et al., 2017b) while the other experiment (Roll et al., 2015), as well as study II (Novén et al., 2021), used very common Swedish nouns (min/max/mean lemmagram frequency of 0.8, 36.7, and 13.90 token counts per million, respectively (Borin et al., 2012)). The participants were given two buttons and instructed to use them to respond as quickly as possible to whether the subject in the sentence got 'one' (singular) or 'many' (plural) things. Apart from using the experiment when collecting EEG data, response times (RT) and ratio of correct responses (accuracy) were recorded. Response times have been used extensively in neuropsychological research at least since F. C. Donders in 1868 (Donders, 1969) suggested that RT relates to neural processing time and gives an idea of the cognitive effort needed to perform a task (Schnikpe and Scrams, 1999; Kyllonen and Zu, 2016). Impact of invalid word accent-suffix combinations could be evident and reflected in both slower and more incorrect responses. Because the word accents are associated with specific suffixes, as discussed in the presentation of Swedish word accents in the Background section, we only compare all accent 1 to all accent 2 stimuli and all valid to all invalid combinations of word accents and suffixes. Direct comparisons of validity within accents are conflated with suffix processing. The difference in RT was thus calculated as $(RT_{a1+sg} + RT_{a2+pl}) - (RT_{a1+pl} + RT_{a2+sg})$. Where RT_{ax+y} mark the mean RT for all items of the accent x and suffix y combinations.

LLAMA tests

This section provides the practical details of the LLAMA language learning aptitude tests used in studies III and IV.

The vocabulary learning aptitude test (LLAMA B) consists of twenty cartoon figures presented simultaneously on a screen. The test taker is given two minutes to click on each figure in any order to make the written names of the figures appear on the screen. After the two minutes, the names of the figures appear one after another and the test-taker is asked to click on the corresponding figure. The score is the percentage of correct responses. Limitations of the test include the fact that the words to memorize are all names of cartoon figures, meaning that only learning of proper nouns is tested. Moreover, because the figures never move, the spatial location on the screen might be used by the test-taker as a mnemonic strategy.

In the phonetic memory test (LLAMA D), participants listen to a series of computer-generated utterances based on names of flowers and natural objects in a British Columbian indigenous language. In the test phase, the test taker is asked to listen to a list of utterances and indicate if they were part of the initial series or not.

In the grammatical inferencing aptitude (LLAMA F) test, the participant is given five minutes to work out the rules governing a foreign language grammar based on twenty written sentences accompanied by cartoon depictions of what the sentences mean. Participants are allowed to take written notes. An important limitation to this test is that it is solely based on the mapping of morphemes and meaning, while aspects of grammar such as variation in word order are not included. This could be a possible addition to the language learning aptitude tests.

The sound-symbol correspondence learning aptitude (LLAMA E) test was not used in the investigations in this thesis. In short, the participant is given two minutes to learn a new alphabet by listening to speech sounds associated with each of 24 letters. After the training phase, the participant is tested by having to click the letter corresponding to a given speech sound until all letters have been tested.

Pitch Perception test

We implemented a computer-based pitch perception test based on the adaptive forced choice pitch direction discrimination task described in (Mathys et al., 2010). The test-takers were asked to indicate if the second of two pure tones played in succession was higher or lower than the first. The tones were played for 400 ms each with a 100 ms pause in between. The first tone was always 550 Hz. If the number of correct answers was ever three more than the number of incorrect answers, the difference in pitch between the tones was reduced. On the other hand, if the number of wrong answers was ever more than the number of correct, the difference was increased. The difference in pitch was initially 30 Hz and then fell through 20, 15, 12.5, 10, 7.5, 6, 5, 4.5, 4, 3.5, 3, 2.5, 2, 1.75, 1.5, 1.25, 1,

0.75 and 0.5 to a minimum of 0.25 Hz. The score was the difference achieved after 100 repetitions of the task.

Participant background tests

Knowledge of the participants' background is important to be able to draw correct conclusions from experimental results. They assist to ensure that the sample of subjects is representative and characterised enough to inform future research developing the studies further, perhaps by applying different selection criteria for subjects. Moreover, the relations between subjects' cognitive skills and their language learning aptitudes help to understand if the aptitudes are specific to the language-learning faculties or reflect more general cognitive skills.

In all studies included in this thesis, participants were university students who grew up monolingual and had no history of mental illness or neurological conditions. We are therefore culpable of adding to the list of studies performed on "Western, educated, industrialized, rich and democratic" (WEIRD) societies, representing about 12 per cent of the world's population (Henrich et al., 2010). It is unknown whether, but improbable that, educational level affects word accent processing.

Handedness

While there are parts of language processing that are bilateral or right lateralised – see e.g. (Hickok and Poeppel, 2007) – there is much evidence that language processing is mainly lateralised to the left hemisphere. Ninety-six per cent of right-handed subjects show left lateralisation of language processing while this is only true for 76 per cent of left-handed subjects (Pujol et al., 1999). As a precaution, only right-handed participants were included in the studies in this dissertation. Handedness was assessed using the Edinburgh Handedness Inventory (Oldfield, 1971) with an inclusion criterion of at least +25 in handedness index. The inventory consists of a short survey in which the participant indicates the preferred hand to use for e.g. drawing or using a knife.

Musicality, working memory, and intelligence

Musicality and musical experience are beneficial for language learning (Wong et al., 2007b; Kraus and Chandrasekaran, 2010; Lee and Lin, 2015). If the language learning aptitude scores strongly correlate with musicality, the interpretations of neuroanatomical correlates of language learning aptitude need to include the possibility that musical experience might have had an impact on brain development. Therefore, musicality and musical experience were assessed using

the Goldsmith musical sophistication index (Müllensiefen et al., 2014). Greater working memory could also potentially explain higher language learning aptitude. If participants are better at keeping information in a short-term storage, this could explain variations in, for instance, vocabulary learning aptitude scores. To address this, we used an automatic version of the operation span test in which participants are tested on how well they remember strings of letters presented while they solve simple arithmetic problems (Unsworth et al., 2005). Greater fluid intelligence, i.e. ability to logically reason, could also have an impact on language learning aptitude, especially on grammatical inferencing aptitude (see relevant parts of the section covering the LLAMA tests). To measure the fluid intelligence of participants, a short version of the Raven's matrices was used (Raven, 2000; Mårtensson and Lövdén, 2011).

Investigations and corrections for these background measures are based on the assumption that language learning aptitude and language proficiency is something unique and possible to separate from e.g. working memory. This might not be strictly true. It is more than possible that the cognitive capacities are part of each other or draw on common neural resources, but for the work presented in this thesis, we are interested in measuring cognitive skills and aptitudes from a linguistic perspective. If we see that the tests are highly influenced by more general cognitive or background measures, we cannot interpret the results as being specific to language capacity.

Magnetic Resonance Imaging

MRI data was used to gather cortical morphometric measurements in all studies as well as tissue microstructure estimates and to segment white-matter tracts in study IV. MRI was chosen because it is safe and non-invasive, thereby not exposing participants to risks as long as proper safety protocols are followed. Moreover, MRI offers uniquely clear contrasts between grey and white matter – as compared to other non-invasive techniques – that make measuring cortical morphometry possible. Furthermore, MRI is currently the only non-invasive imaging technique that allows for diffusion weighting, providing estimates of the structure of tissues on a sub-voxel scale. Possible drawbacks include that it, due to the relaxation effects, is a relatively slow imaging technique, which ultimately limits the obtainable resolution of the images at reasonable participant comfort. Also, we possibly need to exclude participants with metallic implants. However, in the population studied in this dissertation (healthy young adults) these implants are not that commonly present and if they are (for instance screws or plates to fixate broken bones) modern implants are often safe to bring into the scanner room.

For studies II-IV, we chose to use ultra high field (7T) MRI, as compared to 3T in study I, because the increase in static magnetic field means greater contrast to noise-ratio (CNR), thus improving the quality of the cortical morphometry

estimations. CNR is limited by the split in T1 between different tissue types. This split in T1 increases with static field (Korb and Bryant, 2002). Signal strength is important for generating enough signal intensity overall. Signal strength increases with static field strength (Redpath, 1998) but so do inhomogeneities of the field and the impact of susceptibility errors, making quality assurance of fundamental importance (Moser et al., 2012). Higher CNR and SNR can be traded in for higher spatial resolution in a reasonable acquisition time. With a higher resolution, 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). This is argued to be due to less partial volume effects.

Tissue segmentation and cortical morphometry

Typical T1 values for different brain tissue types at 3T and 7T are found in table 1 with values from (a; Marques et al., 2010) or (b; Rooney et al., 2007). This is in no way a comprehensive table but representative.

Table 1: T1-values for brain tissues at 3T and 7T from (Marques et al., 2010) or (Rooney et al., 2007).

Brain tissue type	T1 in 3T (s)	T1 in 7T (s)
White matter ^a	0.810±0.030	1.150±0.060
Highly myelinated grey matter (primary motor cortex) ^a	1.320±0.070	1.870±0.170
Mean cortical grey matter ^a	1.350±0.050	1.920±0.160
Cerebrospinal Fluid ^b	4.427±0.085	4.425±0.137

Intensity in T1-weighted images is inversely correlated with T1, making it possible to segment the tissue types in image volumes. The basis for simple segmentation between cerebrospinal fluid, white matter, and cortical grey matter is fairly straightforward. Assuming only intracranial voxels, if a voxel has a signal intensity that is very low, it is probably cerebrospinal fluid, if it is very bright, it is probably white matter, and sufficiently in-between, grey matter. In voxel-based morphometry, the variations in signal intensities – normalised to some standard – in labelled grey matter voxels are taken as differences in grey matter “concentration”, and grey matter volume is measured as labelled grey matter voxels multiplied with the voxel size (Ashburner and Friston, 1997). However, only studying signal intensities is limiting with regard to registration techniques (Ashburner and Friston, 2000) as well as anatomical metrics, since importantly no information on cortical thickness of surface area is attainable (Dale et al., 1999). The problem of measuring cortical thickness from image volumes directly lies in that the image slices almost never cut the cortex along the geometric normal of the surface, causing an over- or underestimation of the cortical thickness (Fischl and Dale, 2000). An alternative and, as we will see, very useful approach is to compute 3D surfaces corresponding to the white/grey matter boundary (hereafter called the

white surface) and the grey matter/cerebrospinal fluid boundary (henceforth the pial surface). This approach is used in the Freesurfer software suite, containing tools and pipelines for analysing brain anatomical parameters (morphometry). These parameters include the curvature, thickness and surface area of the cortex and automatic labelling of cortical areas as well as subcortical nuclei. After intensity normalisation and skull-stripping (leaving an image volume with only brain tissue voxels with intensity values within useful ranges for the segmentation algorithms) the image volume is preliminarily segmented into grey and white matter by comparing relative intensities while enforcing a planar geometry on the boarder between tissue classes (Dale et al., 1999). Put simply, the neighbourhood of the voxels that are ambiguously classified because their intensity lies in between grey and white matter intensities or because the neighbouring voxels are of another class are analysed to look for a plane that has the most similar intensity. A surface is generated for the white matter volume by tessellating (covering a geometrical plane with surface elements, often triangles, with no gaps or overlaps) the faces of each boarder voxel. Because these surfaces are jagged on the voxel scale, the surface is smoothed using a topology-preserving algorithm. The pial surface is then generated by inflating the white surface until the grey matter/cerebrospinal fluid boundary is found. Finding the pial surface is easier because the intensity difference between grey matter and cerebrospinal fluid is much larger than that between white and grey matter (cf. the T1-values in table 1). Example surfaces overlaid on an MRI image volume can be seen in figure 2. The folding patterns of the cortex can be used to register image volumes to the same space (Fischl et al., 1999) or to automatically parcellate the cortex into anatomical areas (Fischl et al., 2004; Desikan et al., 2006; Destrieux et al., 2010). Cortical surface area is measured as the area of the tessellation triangles (figure 3). Cortical thickness is measured as the shortest distance between each white tessellation triangle and the pial surface and the resulting thickness is saved for each vertex as the average of the tessellation triangles the vertex is part of (Fischl and Dale, 2000). The cortical thickness measurements have been validated and the test-retest reliability has been confirmed (Rosas et al., 2002; Han et al., 2006; Reuter et al., 2012). The surfaces are not restricted to the voxel resolution, so sub-voxel size differences are detectable between groups.

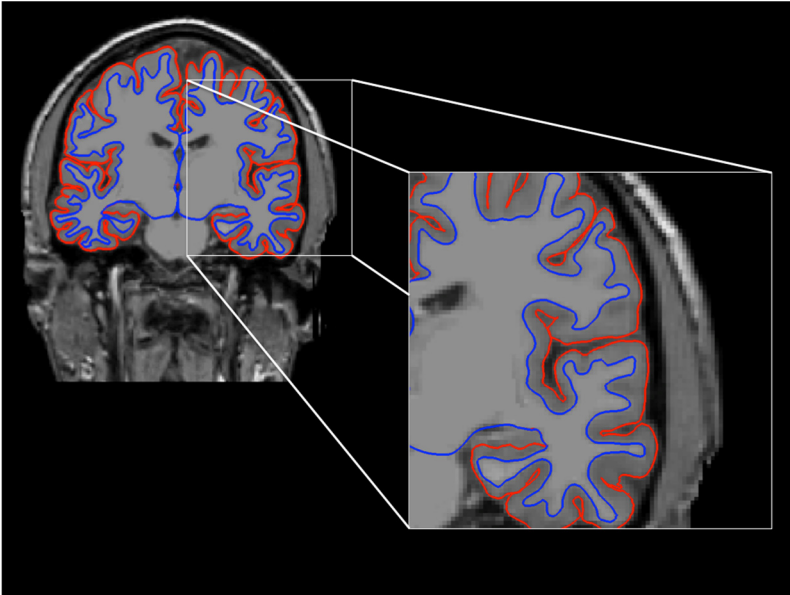


Figure 2: Example FreeSurfer surfaces in a 2D-coronal view. Cortical thickness is measurable as the shortest distance between the border between white and grey matter (blue line) and the pial surface (red line).

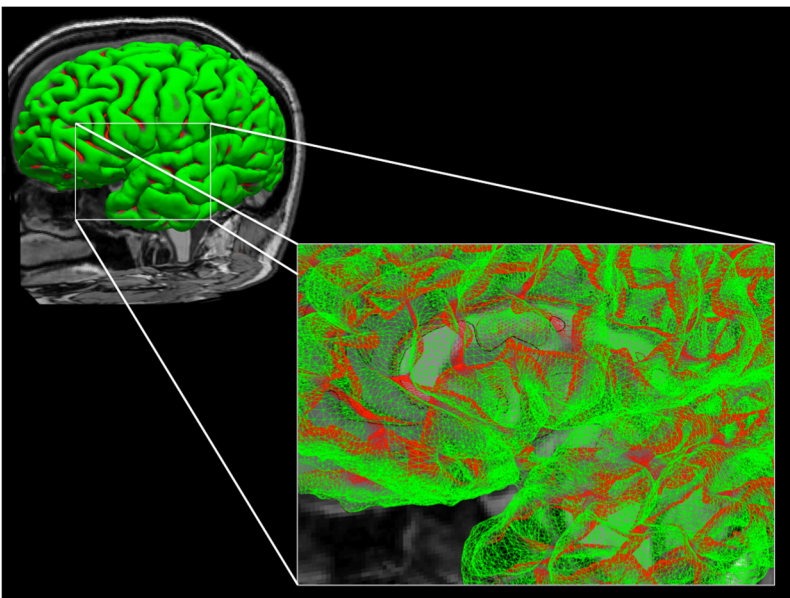


Figure 3: Example FreeSurfer pial surface in a 3D-view. The surfaces are made up of tessellation units, the triangles seen in the zoom-in in the image. These triangles can be used to measure the surface area of the cortex.

dMRI

In study IV we acquired two dMRI volumes. The DKI volume was designed with multiple encoding strengths but lower spatial (2x2x4 mm voxels) and diffusion encoding directional (6, 6, 12, 16 directions) resolution to be able to estimate the diffusion (excess) kurtosis as well as the diffusion tensor. The other, the tractography volume, was more specialised for tract segmentation with a higher spatial (2x2x2 mm voxels) and diffusion encoding directional (56 directions). Diffusion kurtosis adds more information about the tissue microstructure, specifically the complexity and diversity of the water compartments within a voxel, but at the cost of either longer scan time, fewer diffusion encoding directions, or lower spatial resolution to accommodate the additional encoding strengths. This is the reason why the voxel size is different in the DKI than in the tractography volume; we had to assure a reasonable total scan time. Based on experience, ten minutes of active scanning is a sensible time limit before allowing participants to scratch an itch or rest their ears from scanner noise. The tractography sequence had a 9 minute 27 second scan time while the DKI sequence was 4 minutes 40 seconds long.

White matter tract segmentation

Because of ambiguity in the interpretation of the dMRI data on the voxel level it is impossible to know for sure that a proposed connection based on the principal diffusion direction alone is anatomically valid (Jbabdi and Johansen-Berg, 2011). Therefore, we instead used a segmentation tool, TractSeg, to segment well-known anatomical white-matter tracts in study IV. A comparison between the best available tractography algorithms reported that while 90 % of known anatomical fibre bundles were discovered, many more invalid than valid bundles were found (Maier-Hein et al., 2017). It is therefore reasonable to manually intervene in order to only segment known anatomical tracts based on dMRI atlases (Wakana et al., 2004; Catani and Thiebaut de Schotten, 2008). This approach is both laborious and introduces a great deal of subjectivity in the selection of which exclusion criteria to apply. An alternative is to use segmentation software tools that have been trained on large datasets of expert segmentations (O'Donnell and Westin, 2007; Jin et al., 2012; Jin et al., 2014; Wasserthal et al., 2018a). TractSeg is one such tool for automatic white matter bundle segmentation that we used in study IV (Wasserthal et al., 2018a; Wasserthal et al., 2018b; Wasserthal et al., 2019). TractSeg is trained on a cohort from the Human Connectome Project (Van Essen et al., 2013) to segment well-known anatomical tracts directly from fibre orientation distribution functions, obtained by constrained spherical deconvolution (Tournier et al., 2007).

Functional MRI

MRI can also be used to map functional responses, i.e. increases in neuronal activity, to task performance or stimulus presentation. The technique is based on the blood-oxygen-level-dependent (BOLD) contrast (Ogawa et al., 1990; Ogawa et al., 1992), the outline of which is given in the following section. Because the brain does not store oxygen or sugar, this must be delivered via the blood when neurons are active. To conserve energy, blood vessels only open if there is a need for oxygen or energy. When the blood vessels open, more blood flows into the tissue. Oxygenated blood increases the signal strength in T2*-weighted images. The expected increase in signal strength after sufficient neuronal activity is described in a haemodynamic response function (HRF). The typical shape of an HRF is an increase in signal strength reaching its peak after about five seconds and then renormalising, due to the closing of the blood vessel, after an additional twenty seconds. Due to the relatively slow response, the temporal resolution of the technique is not great, but the spatial resolution is its strength. By convolving the time points in which researchers believe the neuronal activity occurs (e.g. when stimuli are presented) with the HRF, a model time course is obtained that can be statistically compared to the measured signal in a series of T2*-weighted images. If the measured signal in a voxel corresponds sufficiently to the model, it is said to be active during the task. In order to measure responses specific to the cognitive processes studied, careful thought must be put into experimental design to compare states of brain activity that only differ in the way that is relevant for what is to be studied.

An fMRI experiment was attempted for study II but the introduced silent gaps for stimulus presentation in the sequence ultimately corrupted the data too much for any significant BOLD contrast to be found.

ERP

In order to relate findings in study II to previous studies, including the ones from which data was used in study I, electroencephalography (EEG) data was recorded. EEG is a technique for measuring electrical activity from the brain using electrodes placed on the scalp. The neural signal component is thought to arise from summed postsynaptic potentials from large numbers (thousands to millions) of similarly oriented neurons, predominantly pyramidal cells, and is on the order of magnitude of 10-100 μV (Vaughan, 1982; Luck, 2012). However, this signal is masked by noise from the electrical equipment as well as from other physiological sources such as eye movements. Such artifacts are identified and corrupt data removed or corrected by using, for instance, independent component analysis algorithms (Jung et al., 2000). The ERP technique is based on measuring EEG on participants while they perform a task or are exposed to stimuli and then averaging

the event-related activity from many such experimental events (Luck, 2005, 2012). The averaging is done to increase the signal and reduce the impact of stochastic (random) noise, i.e. improving the signal-to-noise ratio (SNR). As in fMRI, experiments need to be carefully planned to capture relevant differences in brain states (see relevant parts of the fMRI section above). The resulting average ERP wavelet can then be compared between experimental conditions and conclusions can be drawn on the neural processing involved in the experimental event types.

The temporal resolution of EEG is excellent but the exact positions of the active neurons are impossible to determine. Because the number of possible generators of the signal is unknown and electrodes can only be placed on the top of the head, there are infinite solutions to the so-called inverse problem, i.e. determining the source of a signal measured on a surface. However, scalp distributions showing the recorded signals at different electrodes across the scalp are often provided to hint at the topography of ERP effects.

The components of an ERP are called “negativities” or “positivities”, depending on the direction of their deflections from baseline. Also, specific ERP components are named after their timing relative to stimulus presentation. Accordingly, the N400 component is a negative deflection peaking around 400 ms after stimulus presentation.

Study II included a control ERP experiment to relate findings to previous studies. On the request of reviewers, this experiment was removed from the article but is instead given in the summary of study II in this thesis.

Power analysis

Statistical power refers to the probability of detecting an effect if the effect is truly there, i.e. to correctly reject the null hypothesis (Cohen, 1977). The effect size, alpha level (i.e. the threshold probability for rejecting the null hypothesis), and number of participants are directly related to the power of the analysis. The effect size is in some form a ratio between, in the case of group comparison, the difference between groups or, in the case of correlation, the covariance of the tested quantity and the (pooled) standard deviation. Study IV includes some reasoning about statistical power when relating brain structure to behavioural measures.

The investigations

Study I – Cortical thickness of planum temporale and pars opercularis in native language tone processing

The aim of study I (Schremm et al., 2018) was to test for associations between the cortical thickness and surface area of bilateral planum temporale (PT) and inferior frontal gyrus pars opercularis (IFGpo) and reliance on word accents as cues to word endings. The reliance on word accents as phonological cues was quantified as the difference in RT between valid and invalid Swedish word accent-inflected suffix combinations.

RT and MRI data were taken from two previous studies on Swedish word accent processing, one using real words (Roll et al., 2015) and one using phonotactically legal pseudowords (Söderström et al., 2017b). In both experiments, target word stems were always monosyllabic and carried in inflected forms in sentences of identical structure, e.g. *Kurt fick [target word] till jul*. ‘Kurt got [target word] for Christmas.’ Four variants of each word stem were generated by combining either accent 1 or 2 with either singular or plural suffix. RTs were measured from suffix onset. Participants were asked to, as quickly as possible, indicate if the person in the carrier sentence got “one” or “many” things. Invalid word accent-suffix combinations generated longer response latencies than valid combinations because the word accents are used to predict the upcoming suffix in both the real and pseudoword study. This prediction was also indexed in a PrAN in both studies. That is to say, even if the words carry no lexical meaning, word accents are still used to predict their inflected forms, shown both behaviourally and electrophysiologically. However, the manner in which the predictive processing is performed, and their subsequent possible relationship with cortical anatomy, is different as shown in partially distinct functional activity patterns in Roll et al. (2015) and Söderström et al. (2017).

We could show that the RT difference between valid and invalid word accent-suffix combinations, i.e. the reliance on word accents as phonological cues to suffixes, correlated with cortical thickness of left PT ($r=0.599$, $p=0.30$) for real word processing and with left IFGpo ($r=0.492$, $p=0.045$) for pseudoword processing. Furthermore, cortical surface area was found to correlate negatively with RT difference in the pseudoword experiment in left PT ($r=-0.517$, $p=0.34$) but nowhere else and nowhere for real word RT differences. The fact that RT

differences from real and pseudoword experiments show correlations with cortical thickness in different cortical areas corroborates what is suggested from functional experiments: word accents as phonological cues to word endings are processed differently as part of a lexical word as compared to a pseudoword (Roll et al., 2015; Söderström et al., 2017b). For real words, the importance of the structure of the left PT could be due its role in higher-level acoustic analysis of the speech signal. The thickness of left PT could reflect the efficiency with which the tonal information is extracted and associated with inflected forms in the mental lexicon. Assuming that the inflected forms and word accents are stored as one mental representation, this would be a more efficient processing strategy than laborious rule-based decomposition necessary for the non-lexical pseudowords. Said rule-based processing could instead benefit from a thicker left IFGpo, as suggested by our results, and the benefit from a valid accent cue would then be to preactivate the correct suffix.

Study II – Cortical thickness and surface area of left anterior temporal areas affects processing of phonological cues to morphosyntax

Results from study I showed that the reliance on Swedish word accents as cues to word endings in real words correlates with cortical thickness in the left PT. However, dual stream models of neural speech processing, described in detail in the Background section, led us to hypothesise that more ventral speech processing areas might also be of importance for making use of word accents as cues to word endings as word accents facilitate both the lexical semantic (i.e. the meaning) and morphosyntactic (in this context, the inflected form) processing of a word. In fact, even if PT is mostly considered part of the dorsal stream, it partially covers the ventral stream as well (DeWitt and Rauschecker, 2013). This motivated us to apply a spatially unbiased whole-brain analysis to investigate the impact of cortical thickness and surface area on RT differences between valid and invalid word accent-suffix combinations. Moreover, we used an MRI scanner with higher field strength (7T) to benefit from larger tissue contrasts. The behavioural task was the same as in study I, but adapted for an fMRI setting with longer inter-stimulus intervals and stimuli with a slower speech rate.

Results showed that cortical thickness as well as surface area in the left anterior superior temporal sulcus and middle temporal gyrus correlated with perceptual phonological proficiency (cortical thickness: $t(12)=3.15$, Pearson's $r=0.563$, $p=0.00010$; cortical surface area L. middle temporal gyrus: $t(12)=7.46$, Pearson's $r=0.598$, $p=0.00010$; cortical surface area L. inferior temporal gyrus: $t(12)=5.58$, Pearson's $r=0.598$, $p=0.024$). This implies that perceptual phonological proficiency can possibly manifest in the cortical morphology of the ventral speech

processing stream areas. The reliance on word accents as cues to word endings was found as faster ($t(23) = 3.21, p = 0.0020$) and more accurate ($t(23) = 3.08, p = 0.00263$) responses for correct than incorrect word accent-word ending combinations. The original experiment also included an EEG experiment presented at the end of this summary, but it was not included in the article following the peer review process. In short, the EEG study replicated the results from previous ERP investigations on word accent processing and showed that the participants used the word accents to access morphosyntactic and semantic information.

EEG control experiment in Study II

Introduction

In order to compare the results in study II with previous EEG findings, a control experiment was conducted shortly after each participant's MRI session.

The greater predictive value for accent 1 than accent 2 is reflected in an early event-related potential (ERP) known as pre-activation negativity (PrAN) for accent 1 relative to accent 2 that correlates with the predictive value of the accent on a word stem (Söderström et al., 2016) as well as shorter response times for validly cued accent 1 suffixes than for accent 2 suffixes (Söderström et al., 2012). Importantly, the predictive value of word accents in Central Swedish is preserved in a dialect where the acoustic realizations of the accents are reversed (Roll, 2015). Moreover, ERP studies have shown that invalid accent cues for suffixes in Swedish elicit either a left anterior negativity (LAN; Roll, 2015; Söderström et al., 2017a) or no negativity (Roll et al., 2010; Roll et al., 2013; Roll et al., 2015) before a P600. Left anterior negativity (LAN) is elicited by morphosyntactic errors, possibly reflecting increased morphological processing (Penke et al., 1997; Weyerts et al., 1997). P600 is elicited by syntactic errors and ambiguities and correlates with processing of syntactically complex sentences and can be interpreted as a late stage of reanalysis (Friederici, 2002). Söderström et al. (2017) also found an increased N400 for invalidly cued suffixes in pseudowords carrying stem-tones. Since N400 increases in amplitude if a word cannot be integrated semantically into the preceding context (Kutas et al., 2006; Kutas and Federmeier, 2010), the increased N400 is thus understood, together with the P600, to reflect processing cost of failed predictions from tonal cues as to which suffix is to follow.

Methods

The stimuli and task were the same as described in Novén et al. (2021), but in the EEG experiment, every stimulus presentation and response recording was 4000 ms long followed by a silence. The duration of the silence was jittered among the stimuli between 100-1000 ms.

The ERP experiment was performed using Eprime software (2.0.10.353). A 32 channel system from BrainProducts (Brain Products GmbH) was used. Electrode impedances were kept under 5 k Ω . Activity was sampled at 5 kHz. Recorded data were re-referenced to average mastoids off-line. Data were high-pass filtered at 0.01 Hz online and low-pass filtered at 30 Hz offline. Independent component analysis was performed to remove ocular artefacts. Baseline correction was made based on a 200 ms time-window prior to the time-locking point. ERP data was analysed in 1 s epochs following accent and suffix onset. Three time windows were created based on previous literature and visual inspection of the data. Electrodes were grouped into four regions of interest: left anterior (F7,F3), left posterior (P7,P3), right frontal (F8,F4), and right posterior (P8,P4). Mean ERP amplitudes were submitted to a repeated measure Analysis of Variance (ANOVA) with the within-subject factors validity (Valid vs Invalid) or Accent (1 vs 2), anteriority (anterior vs posterior), and hemisphere (left vs right). Data from 3 participants had to be discarded due to problems with data registration.

Results

Behaviour

An upper-tailed paired-sample t-test revealed significantly greater response times for invalidly ($M=880$ ms, $SD=250$) than validly ($M=827$ ms, $SD=236$) cued suffixes $t(23)=4.45$, $p=9.11e-05$. While disturbed by invalid accent-suffix association, participants could still perform well as shown by a non-significant decrease in accuracy ($M_{\text{valid}}=0.990$, $SD_{\text{valid}}=0.0118$; $M_{\text{invalid}}=0.984$, $SD_{\text{invalid}}=0.0212$; $t(23)=1.30$, $p=0.208$). Response times were significantly shorter for validly cued singular – accent 1 – than validly cued plural – accent 2 – stimuli $t(23)=-2.87$, $p=0.00867$, lower-tailed paired-sample. There was no significant difference in retard effect due to invalidly cued suffix, i.e. response time difference between invalidly and validly cued suffixes, between accents ($M_{\text{accent1}}=64.8$ ms, $SD_{\text{accent1}}=75.2$; $M_{\text{accent2}}=41.6$ ms, $SD_{\text{accent2}}=76.1$; $t(23)=1.19$, $p=0.248$). These results indicate that word accents were used to facilitate processing of grammatical number suffixes by the participants to a similar degree.

Word accent effects

Accent 1 yielded increased negativity compared to accent 2 stimuli (figure 4) 200-300 ms after tone onset in the target words as shown by an Accent \times Anteriority interaction $F(1, 20)=5.07$, $p=0.0357$. This is interpreted as a PrAN. Follow up ANOVAs showed main effects of accent at anterior $F(1,20)=13.8$, $p=0.00136$, but not posterior $F(1,20)=0.254$, $p=0.619$ electrodes.

Response time difference between accent 1 and accent 2 from the fMRI – but not EEG – experiment correlate with mean amplitude between 200 and 300 ms from the left anterior ROI ($r(28)=0.365$, $p=0.0473$).

Suffix effects

Invalidly cued suffixes elicited increased negativity 275 to 350 ms after suffix onset interpreted as a LAN (figure 4) as shown by a significant Validity \times Hemisphere interaction, $F(1,20)= 4.75, p= 0.0413$. Follow-up ANOVAs showed a main effect of Validity in the left, $F(1,20)= 16.2, p= 0.000671$, but not right, $F(1,20)= 1.82, p= 0.192$, hemisphere. Both hemispheres showed a main effect of anteriority, left: $F(1,20)= 21.9, p= 0.000145$ right: $F(1,20)= 12.0, p= 0.00241$.

Invalidly cued suffixes elicited a principally posterior positivity at 500-700 ms as shown in a main effect of validity, $F(1,20) = 4.87, p= 0.0392$, and anteriority $F(1,20) = 33.8, p= 1.10e-05$. This is interpreted as a P600 effect of invalidly cued suffixes.

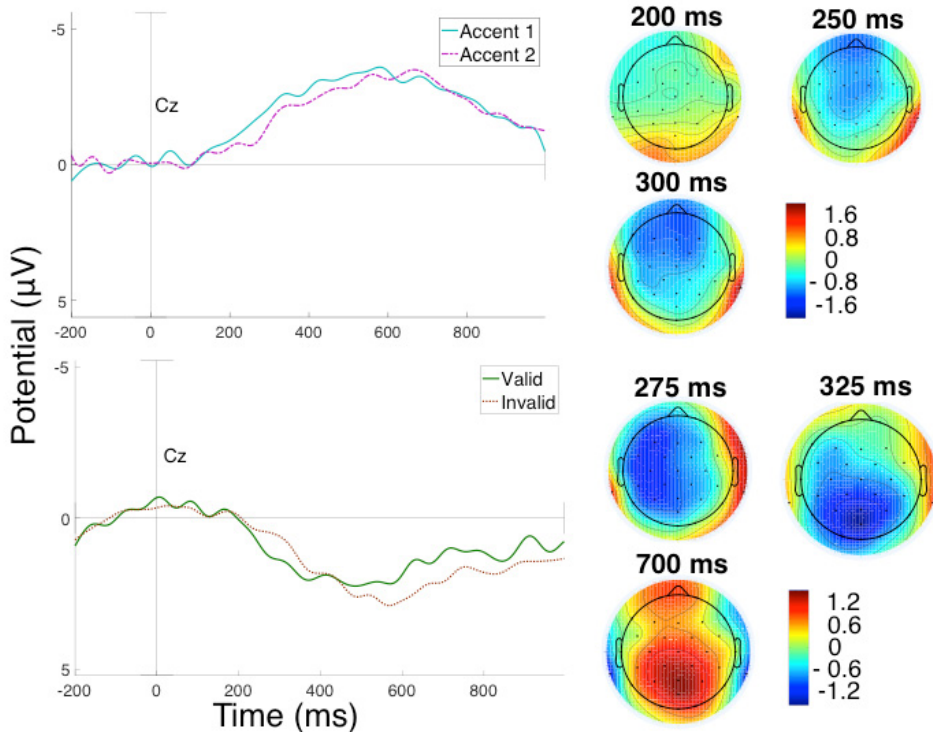


Figure 4: Top left: ERP of accent 1 and accent 2 stimuli time-locked to suffix onset from Cz electrode. Top right: PrAN topography with times from accent onset. Bottom left: ERP of validly and invalidly cued suffixes time-locked to suffix onset from Cz electrode. ERPs were low-pass filtered at 15Hz for display purposes. Bottom right: LAN and P600 topographies with times from suffix onset.

Comparison of fMRI and ERP experiments

The difference in response times between mismatched and matched stimuli did not differ significantly between fMRI ($M=77.2$ ms, $SD=118$) and EEG ($M=58.1$ ms, $SD=57.7$) measurements, as shown in a paired-sample t-test $t(23)=-0.911$, $p=0.372$; neither did the differences in accuracy (fMRI: $M=0.0215$, $SD=0.0342$, EEG: $M=0.00563$, $SD=0.0213$, paired t-test: $t(23)=-1.70$, $p=0.102$). However, the RT differences from the fMRI and EEG experiments did not correlate ($r(22)=0.0483$, $p=0.823$). All participants improved significantly in accuracy in the EEG experiment compared to the fMRI experiment (paired t-test $t(23)=5.73$, $p=7.86e-06$, mean difference = 0.0413). Participants were significantly faster at the task in the EEG than in the fMRI experiment (mean difference: 2010 ms, $t(23)=27.4$, $p<2.20e-16$, paired t-test).

Discussion

Expected ERPs from word accent-suffix processing were replicated from previous studies, and also confirmed that the processing dynamics captured in the ERPs are robust despite previous exposure to the stimuli in the fMRI experiment.

ERP results indicated that participants processed the word accents as phonological cues to suffixes in a manner similar to that observed in previous studies. Moreover, the PrAN and P600 effects evoked from word accents were observed to be robust over time as the EEG experiment took place after the fMRI experiment. This robustness corroborates the idea that early effects like PrAN are automatic effects (Hahne and Friederici, 2002). The negativity elicited by accent 1 word stems is interpreted as a PrAN while the negativity and succeeding positivity following invalidly cued suffix onsets are interpreted as a LAN and P600, respectively. The PrAN was slightly later than in a majority of previous studies. This is most likely due to a slower speech rate ($t(29)=-18.1$, $p=2.20e-16$, $M_{diff}=-0.341$ syllables per second) in the stimuli in this compared to those in Roll et al. (2015; 2017). The LAN effect is associated with morphosyntactic violations and might be involved in activating unprimed suffixes (Pulvermüller and Shtyrov, 2003). The broad distribution of the LAN across the left hemisphere might be due to involvement of more central and posterior cortical areas as the prediction error is based on mismatches involving both phonology and morphosyntax. The P600 likely reflects a reanalysis of the word form as presumptions about the morphosyntax of the word are contradicted by the occurrence of the non-anticipated suffix. Overall, the ERP results replicate previous findings and show that participants use word accents to access morphosyntactic and semantic information.

Study III – Cortical thickness of Broca’s area and right homologue is related to grammar learning aptitude and pitch discrimination proficiency

Studies I and II present results suggesting that the cortical morphometry in brain areas relevant for speech processing correlates with perceptual phonological proficiency in native speakers of Swedish. One way to interpret these results would be to say that people with a thicker cortex or larger cortical surface area have an aptitude for utilising word accents as cues to morphosyntax in their speech processing. This led us to hypothesise that other language aptitudes might be reflected in the regional thickness of the cortex with a thicker cortex indicating greater neural processing power. This was investigated in study III (Novén et al., 2019).

The LLAMA test battery (described in detail in the Background) was used to measure participants’ language learning aptitude. As we wished to relate our findings to the results in studies I-II concerning tonal contrasts, we chose to also measure the participants’ ability to perceive pitch. We could show that cortical thickness in left inferior frontal gyrus pars triangularis (IFGpt), part of Broca’s area, as well as in left medial superior frontal gyrus correlate with aptitude for inferring a new grammatical system, LLAMA F score, (IFGpt: $r(30)=0.65$, $p=0.0202$; medial superior frontal gyrus: $r(30)=0.66$, $p=0.0137$) while the cortical thickness of the right-hemispheric Broca homologue correlates negatively with pitch perception ($r(30)=-0.57$, $p=0.0006$). No correlations were found for the LLAMA B or D subtests.

The LLAMA F test consists of inferring the meaning of morphemes from pairs of written sentences in a foreign language and cartoon figures described by the sentences. Participants who score highly on the LLAMA F test had thicker cortices in the left medial superior frontal gyrus, known to be involved in deductive reasoning (Goel et al., 1997) as well as in left IFGpt. The left IFG is crucial for language use and pars triangularis is known to be involved in processing combinations of lexical elements on the word level, rather than syntactic processing (Hagoort, 2005; Goucha and Friederici, 2015). Taken together, grammatical inferencing aptitude is correlated with cortical thickness in cortical areas associated with deductive reasoning, in this context beneficial for the process of determining which morphemes mean what, and merging the meanings of morphemes, and possibly advantageous for understanding how the morphemes relate to one another. Pitch perception proficiency, on the other hand, was shown to correlate negatively with cortical thickness in the right IFG. Although most studies investigating the associations between cognitive performance and cortical thickness find positive correlations, studies on people on the extreme points of the spectrum of pitch perception ability have shown an interesting pattern. Musicians with absolute pitch (i.e. the ability to identify or re-create a musical note without a

reference tone) have thinner cortex in, among other areas, the right IFG (Bermudez et al., 2009) while those suffering from congenital amusia (i.e. tone deafness) have a thicker right IFG (Hyde et al., 2007). Our findings fit this pattern of “more is worse” for pitch perception well.

Study IV – Cortical and white matter correlates of language learning aptitudes

For study IV we wanted to investigate the brain anatomical correlates of the LLAMA tests further and thus we used T1-weighted images with higher resolution and diffusion-weighted MRI to investigate not only the cortical thickness and surface area but also the diffusion properties of known language-associated white-matter tracts. This study also included more participants (54 as compared to 32 in study III). Mean values of the DKI parameters were extracted from white matter tracts well-known to be language relevant, i.e. the arcuate fasciculus (AF), subsegment III of the superior longitudinal fasciculus (SLF III), the uncinate fasciculus (UF), and the inferior frontal-occipital fasciculus (IFOF). Unfortunately, the parameter values from UF and IFOF deviated too much from expected values to be used in further analyses. This was probably due to imaging artefacts that could possibly be mitigated in future studies by use of lower field-strength MRI or better correction using magnetic field maps.

We were able to show that cortical surface area in a left posterior medial cluster correlated with vocabulary learning aptitude (LLAMA B) score ($r(54) = 0.485$, $p = 0.00840$). The cluster covers the inferior part of the precuneus and has been shown to be part of hippocampal-cortical memory systems (Vincent et al., 2006; Kahn et al., 2008; Vincent et al., 2008) as well as integration of visuo-spatial information with declarative memory (Cavanna and Trimble, 2006). This indicates that a greater cortical surface area is beneficial for storing written word-figure associations. It would be interesting to investigate in future studies if the correlation between cortical surface area in the left inferior precuneus with vocabulary learning score is due to the visual nature of the LLAMA B test. That is, do the results carry over if the vocabulary-learning task is, for example, auditory instead? Moreover, improvements in the dMRI sequences should be made to also allow for investigating the correlations with the microstructure of UF as it is connected to the hippocampus (Saur et al., 2008) and thereby to memory encoding and retrieval.

Phonetic memory (LLAMA D) score instead correlated negatively with axial kurtosis in the left AF and SLF III (AF: $r = -0.523$, $p = 5.92e-5$; SLF III: $r = -0.458$, $p = 5.70e-4$). This can be interpreted as meaning that phonetic memory benefits from less tissue complexity along the fibres. Left AF and SLF III connect cortical areas that have been shown to be important both for learning novel speech sounds

(Golestani et al., 2002; Golestani et al., 2007; Veroude et al., 2010; Golestani et al., 2011) as well as for phonological working memory (Strand et al., 2008; McGettigan et al., 2010; Scott and Perrachione, 2019). LLAMA D is an implicit learning task and the structure of the left AF and SLF III might, speculatively, indicate that there is a more active perception and memory network for speech sounds in persons with better phonetic memory.

We were unable to replicate the findings in study III that cortical thickness in IFGpt correlates with grammatical inferencing aptitude (LLAMA F) score. This could be due to the higher resolution in this study ($0.8 \times 0.8 \times 0.8 \text{ mm}^3$ as compared to $1 \times 1 \times 1 \text{ mm}^3$) that could cause less variability in the cortical thickness estimations. Additionally, the greater number of subjects in this study could mean that there is a greater variance in the strategies applied to perform the task, including strategies that do not benefit from a thicker IFGpt. This could be investigated further by refining the tests to more closely match specific processing roles in focal brain areas.

There are potential limitations in statistical power in this and other studies of associations between brain structure and behavioural measures. Reasonable estimates of effect sizes of correlations between white matter microstructure and language-related proficiencies range from $r=0.35$ to 0.56 and for cortical morphometry between $r=0.46$ to 0.66 . If we want to apply a significance threshold of $p=0.05$, Bonferroni corrected for comparisons between 4 ROIs for diffusion parameters and 58 ROIs for cortical morphometry, and a power of 0.8 , the number of subjects needed to detect the effect range from 23 to 86. The number of ROIs for cortical morphometry are taken from the cortical parcellation included in FreeSurfer (Desikan et al., 2006). This means that while the reported results are not at immediate risk of being underpowered, future studies can benefit from including more subjects, at least 86.

General discussion

The four studies presented in this dissertation show that proficiency in certain aspects of language use correlates with a thicker cortex, larger cortical surface area, and more ordered white matter tracts. Assuming that a thicker cortex means greater number of cells or dendritic branches within functional columns (Rakic, 1988; Zatorre et al., 2012), it would mean that participants with greater proficiency have, in a sense, more processing power in that they can take input from more cells (from more dendritic branches) and transmit more signals from the cellular calculations (from more neurons) or get more support from more cortical glial cells. Greater cortical surface area would instead mean more functionally specialized cortical columns (Rakic, 1988; Rakic, 2000) and thus the possibility of richer representations or higher resolution in the areas' functional roles. Language processing relies on efficient signalling between functionally specialised cortical

areas through white matter tracts. DKI parameters offer an insight into the tissue microstructure of such white matter tracts and thus offer some measure of quality of the connections between cortical areas. Taken together, different language proficiencies benefit from different configurations of cortical and white matter tract anatomy in ways that are detectable using MRI.

Results from studies I and II show that degree of perceptual phonological proficiency in native speakers of Swedish, measured as extent of reliance on word accent cues in predicting word endings, is associated with a thicker cortex and larger cortical surface area in temporal, predominantly ventral speech processing stream areas for real words. The association between cortical morphometry in areas in or adjacent to this gradient indicate that the perceptual phonological proficiency manifestation is related to both lexical access and combinatorial processing of lexical elements. The results relate to findings in fMRI experiments that have highlighted left PT, HG, and adjacent superior temporal gyrus as a source of the word accent related pre-activation of suffixes (Roll et al., 2015; Söderström et al., 2017b). Both the results of study I and Söderström et al. (2017b) indicate that the role of the temporal areas is more pronounced for lexical word stems than for pseudoword stems as regards using word accents as predictive cues to inflected suffixes. This is nuanced by the findings reported in study II, where expanding the analysis to cover the whole brain instead showed an association between perceptual phonological proficiency and cortical morphometry of the left mid- and inferior temporal gyri. These areas have been shown to be more active when processing the more predictive accent 1 than accent 2 (Roll et al., 2015). Taken together, a speculative interpretation is that greater activity and thicker/larger cortex in anterior middle and inferior temporal gyrus reflect the greater predictive value of and proficiency in making use of the phonological cue, respectively.

To discern the details of the functional role of these areas with regard to perceptual phonological proficiency, neuroimaging techniques with a higher temporal resolution with good spatial localisability such as magnetoencephalography (MEG) or implanted electrodes could be used. An alternative approach would be to induce so-called ‘virtual lesions’ using repeated transcranial magnetic stimulation (rTMS) that is used more and more to study the functional role of cortical areas (Hartwigsen, 2015). The basic idea behind rTMS is to perturb the electrical activity of the brain tissue by applying electromagnetic fields to the area. It is mostly effective on the crowns of the gyri as those are closest to the rTMS coils. The middle temporal gyrus would thus be an interesting candidate, based on the findings in study II, to investigate as regards its exact role in perceptual phonological proficiency in Swedish.

Conflicting word accent-suffix combinations induce greater activity in the left inferior parietal lobe and bilateral supplementary motor areas and middle frontal gyri (Roll et al., 2015; Söderström et al., 2017b). This has been interpreted as being due to reanalysis of the grammatical number (singular or plural) necessary

after failed prediction of which suffix should end the word (Roll et al., 2015). There is no reason why perceptual phonological proficiency would affect the cortical morphometry of these areas as the step of reanalysis comes after the processing of word accents as phonological cues to inflected suffixes, hence the superficially surprising lack of correspondence between correlation between cortical morphometry and response time difference between invalid and valid accent-suffix combinations and activity induced by invalid over valid accent-suffix combinations.

Study II consisted of both an fMRI and an (unpublished control) EEG experiment and this allowed for comparisons in the response times and accuracies between the two experimental settings. These indicate that the MRI scanner environment and/or the novelty of the experiment task increased the dependency on word accents as predictive cues. This is deduced from the following observations: First, task accuracy was weaker in the fMRI than in the EEG experiment, where accuracy was near perfect. Second, accuracy was significantly lower for invalidly than validly cued suffixes in the fMRI but not in the EEG experiment. Third, accent 1 led to a larger difference in response times between invalidly and validly cued suffixes than accent 2 solely in the fMRI experiment. This is expected if participants depend more on phonological cues such as accent 1 being associated with a well-defined set of word endings in contrast to accent 2 (Roll et al., 2010; Roll et al., 2015; Söderström et al., 2016). There is likely a spread in the dependency on cues from different types and modalities. Such a spread is indicated by response time advantage for validly cued suffixes not differing significantly in the EEG and fMRI experiments, but also not correlating. Taken together, we hypothesize that participants who are more sensitive to word accents as phonological cues to morphosyntax have an advantage when it comes to speech perception in noisy contexts and this hypothesis should be investigated in future studies.

Studies III and IV show that language learning aptitude can correlate with cortical morphometry and tissue microstructure in language-related white matter tracts. Grammatical inferencing aptitude could benefit from thicker left inferior frontal gyrus pars triangularis and medial frontal gyrus. This might, however, only hold for some strategies for solving the task. Greater cortical surface area in left inferior precuneus is associated with greater aptitude for learning written vocabulary, which might be due to better mapping of associations between written words and figures in a declarative memory system. Phonetic memory correlates with a more coherent and simple left AF and SLF III that could reflect a more active implicit phonological memory system. Improvements in the dMRI sequences are needed in order to mitigate the signal loss in the inferior parts of the brain in order to examine the roles of the UF and IFOF for language learning aptitudes.

We found in study III that cortical thickness in left IFGpt correlates positively with LLAMA F score while the thickness in the right IFGpt correlates negatively

with pitch perception. This led us to speculate that tasks that are cognitively complex, such as the LLAMA F test, benefit from a thicker cortex more capable of handling variations in input as well as association between different kinds of information, while cognitively simpler tasks, such as the pitch perception test, might benefit from a more streamlined, thinner cortex. This is a suitable hypothesis for future studies, possibly on large open datasets.

Non-linguistic pitch perception correlates negatively with cortical thickness in the right IFGpt while perceptual phonological proficiency correlates with cortical thickness and surface area in left temporal areas. This adds to the evidence that pitch patterns, when having more linguistic importance, draw on neural substrates in the left hemisphere for being processed (Lee et al., 2017).

Looking at the contents of the LLAMA (and other language learning aptitude) test battery in relation to the role of tones in languages (Swedish word accents but perhaps more importantly grammatical tones in e.g. Somali and lexical tones in e.g. Mandarin) it is evident that something might be lacking. Pitch perception has been shown to be a better measure of lexical tone learning than musicality or measures of general language learning aptitudes (Bowles et al., 2016). Some measure of pitch perception should be validated with regards to its value as a predictor for successful lexical or grammatical tone learning and included in, at least, a language learning aptitude test battery for tonal languages.

Another possible addition to language learning aptitude tests would be some measure of sensitivity to linguistically relevant structures of information (i.e. abstract syntactic structures). The LLAMA F test has more to do with agreement than with syntax, as word order is basically not given importance. A viable candidate for a more syntax-oriented test would be one based on artificial grammar learning (AGL), first introduced by Arthur Reber (1967). The standard application of the AGL paradigm is to ask participants to remember strings of letters without telling them that they are derived from a set of rules. These rules make some combinations of letter sequences "illegal". After the training phase, the participants are told that there is a set of rules governing the generation of the letter strings. They are then asked to judge the grammaticality of new letter strings. While AGL has been argued to be most applicable for understanding first language acquisition, it could also be a relevant tool for studying implicit aspects of adult language learning (Uddén and Männel, 2018). AGL and incidental language learning, i.e. learning without instruction, are related but different processes, possibly indicating that the standard AGLs do not reflect processes in the initial stages of foreign language learning (Robinson, 2010). The neural basis for AGL has been studied and seems to be both similar and different to natural language processing, for a review see (Uddén and Männel, 2018). In sum, AGL could possibly serve as a basis for a language aptitude test but would then have to be adopted into a version that robustly predicts foreign language learning success. It would then complement the grammatical inferencing skill with a measure of (implicit) statistical learning of linguistically relevant structures.

Conclusions

This dissertation presents four studies that have all investigated neuroanatomical correlates of language ability, either perceptual phonological proficiency or language learning aptitude. We have found that variations in cortical thickness, cortical surface area, and white matter microstructure are indeed associated with differences in language proficiency. Thereby, we add evidence to the general concept that language learning aptitude and proficiency can manifest as differences in brain structure.

Cortical morphometry in temporal speech processing areas is associated with Swedish phonological proficiency

Study I reported that perceptual phonological proficiency, measured as impact of mismatched word accent-suffix combination on response time, correlated with cortical thickness in left planum temporale or left inferior frontal gyrus pars opercularis, depending on whether the word stems were part of lexical words or pseudowords. This could indicate that frequent lexical words are stored together with accent and suffix in the mental lexicon in subjects with high perceptual phonological proficiency and that this storage benefits from a thicker cortex. On the other hand, for pseudowords, where no whole-word storage is possible, mismatch in word accent-suffix combinations still led to slower responses. The correlation between greater cortical thickness in left inferior frontal gyrus pars opercularis and perceptual phonological proficiency points towards a greater accommodation of word accent-suffix associations in rule-based, decompositional analysis. In study II, the investigation was broadened to cover the entire cortex and perceptual phonological proficiency was found to correlate with cortical thickness and surface area in left middle and inferior temporal gyrus, parts of the ventral speech processing stream. The ventral processing stream maps sounds to meaning and the literature shows that, in fact, part of planum temporale is part of the ventral stream as well. Taken together, results from studies I and II show that greater cortical thickness and surface area in left temporal areas correlate with a greater ability for using phonological cues to word endings in real words.

Brain structure correlates with language learning aptitude

Looking instead at neuroanatomical correlates of language learning aptitude, we could show in study III that cortical thickness in left inferior frontal gyrus pars triangularis correlates with grammatical inferencing ability, although this could not be replicated in study IV using more participants and MR images of higher resolution. Study IV could instead show that vocabulary learning aptitude score is associated with greater cortical surface area in left inferior precuneus, suggesting greater memory encoding ability for written word-figure associations. Moreover, axial diffusional kurtosis in left arcuate fasciculus and subsegment III of the left superior longitudinal fasciculus correlated negatively with phonetic memory aptitude, interpretable as more coherent, less complex tissue in dorsal stream-related white matter tracts being beneficial for phonetic working memory.

While grammar inferencing aptitude score is positively correlated with cortical thickness in left inferior frontal gyrus pars triangularis, skill in the cognitively simpler pitch perception task profits instead from a relatively thinner cortex in its right-hemisphere homologue. This could speculatively implicate that simpler tasks benefit from a more efficient and specialised cortex while more complex tasks instead benefit from a thicker cortex, capable of creating and maintaining associations between different kinds of information.

Outstanding issues and future directions

Word accents

Results from studies I and II as well as previous work on word accent processing provide a groundwork for studying new aspects of word accents as phonological cues to word endings. For instance, if it could be known when Swedish word accents are integrated as phonological cues to word endings, correlations between the development of perceptual phonological proficiency and the brain could provide insights into the nature and impact of this aspect of language acquisition. Studies have shown that children cannot use the phonological alternation of the determiner *a/an* as phonological cues in English until at least more than five years of age (Gambi et al., 2018) and that pitch accent cues for compounds in Japanese are used from at least six years of age, the exact age of acquisition being unknown (Hirose and Mazuka, 2017). In other words, although the children are aware of the contrast, they do not use it predictively. If the same holds for Swedish native speakers, children from five years of age could be investigated to find those proficient at using phonological cues. The hypothesis, based on the findings in studies I and II would be that the cortex in left anterior-temporal areas would develop to become relatively thicker and larger in terms of surface area for the more proficient users of phonological cues to morphosyntax. The fact that left superior temporal and inferior frontal areas increase in thickness in children between the ages of 5 to 11, an age also associated with regional cortical thinning, (Sowell et al., 2004) speaks in favour of the left anterior-temporal areas being plastic during a time when word accents are given importance as phonological cues. However, the finding in study I that the integration of word accents as cues to suffixes in more rule-based processing of pseudowords might indicate that the association between word accent and full wordform has to be well-established and therefore an initial importance of the thickness in left inferior frontal gyrus might be expected at first.

Language learning aptitude

Studies III and IV show correlations with brain anatomy and language learning aptitude that help inform on which brain structures are important in what ways for having a greater aptitude for language learning. However, the available language learning aptitude tests are developed to detect talented language learners, not to pinpoint the possible neural contributions for the aptitude to diversify. This means that future studies should attempt to further divide the components of language learning aptitude tests into subcomponents until obtaining a granularity that matches with what the mechanisms within the brain do. This is a reflection of the brain-cognitive behaviour problem presented in (Buzsáki, 2020). He argues from the standpoint of a “radical implementationist”, arguing that all studies in cognitive neuroscience should define descriptors of behaviour from brain mechanisms, not burdened by philosophical connotations. However, I agree with the important modifying counterpoint by (Poeppel and Adolfi, 2020) stating that descriptors of behaviour from e.g. psychology or linguistics can in principle be broken down into meaningful descriptors.

Moreover, language learning aptitude tests are possibly limited by not testing the sensitivity for novel phonological contrasts or for inferencing syntactic properties in a novel language. While the LLAMA tests include a phonetic memory component, aptitude for using phonological contrasts in novel ways such as for making semantic distinctions are not tested. Aptitude for perceiving and using these contrasts are important for learning tonal languages and the results presented in this dissertation suggest that linguistically relevant pitch perception benefits from greater cortical thickness and surface area in left temporal areas as well as cortical thickness in the left inferior frontal gyrus. Grammatical inferencing skill and the corresponding parts of the modern language aptitude test (grammatical sensitivity and inductive learning ability) test more the ability to infer the meaning of words and morphemes and not much emphasis is placed on the importance of word or morpheme order or morphological inflections. A starting point for such a subtest could be artificial grammars widely used in experiments to study acquisition of syntax (Uddén and Männel, 2018) while being aware of their potential limitations (Robinson, 2010).

Are thicker cortices better for high-level cognition but thinner better for low-level cognition?

Results in study III indicate that performance in tasks with a low cognitive load might benefit from thinner, more efficient, cortex while performance in tasks with a high load might benefit from a thicker cortex. This should be investigated in future experiments. Previous studies have found that thinner cortex is sometimes

associated with better cognitive performance in children (Schnack et al., 2015; Botdorf and Riggins, 2018). However, this likely reflects different stages of cortical development as the cortex thins during maturation (Sowell et al., 2004; Tamnes et al., 2011). This apparent thinning might be due to increased myelination of the cortex, making it look more like white-matter in T1-weighted MRI images, at least in primary sensory areas (Natu et al., 2019). One possible contradiction would be that the association between general cognitive abilities and cortical thickness and surface area is complex, with positive and negative correlations across the cortex (Vuoksimaa et al., 2016). However, the measure of general cognitive ability used in the cited study is multi-faceted. Relating back to the reasoning about granularity of the behavioural tasks in the section above about granularity, perhaps it would be possible to divide the cognitive tasks into their less complex subcomponents and compare the morphometrical correlations to skill in both the constituent tasks and the full complex task. To the best of our knowledge, study III and this dissertation are the first to state an explicit hypothesis regarding the link between cognitive complexity and benefit of a thicker or thinner cortex.

References

- Ashburner J, Friston K (1997) Multimodal image coregistration and partitioning--a unified framework. *Neuroimage* 6:209-217.
- Ashburner J, Friston KJ (2000) Voxel-Based Morphometry—The Methods. *NeuroImage* 11:805-821.
- Basser PJ, Mattiello J, LeBihan D (1994a) Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. *Journal of Magnetic Resonance, Series B* 103:247-254.
- Basser PJ, Mattiello J, LeBihan D (1994b) MR diffusion tensor spectroscopy and imaging. *Biophys J* 66:259-267.
- Basser PJ, Pajevic S, Pierpaoli C, Duda J, Aldroubi A (2000) In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine* 44:625-632.
- Benner J, Wengenroth M, Reinhardt J, Stippich C, Schneider P, Blatow M (2017) Prevalence and function of Heschl's gyrus morphotypes in musicians. *Brain structure & function* 222:3587-3603.
- Bermudez P, Lerch JP, Evans AC, Zatorre RJ (2009) Neuroanatomical Correlates of Musicianship as Revealed by Cortical Thickness and Voxel-Based Morphometry. *Cerebral Cortex* 19:1583-1596.
- Borin L, Forsberg M, Roxendal J (2012) Korp – the corpus infrastructure of Språkbanken. In: LREC, pp 474-478. Istanbul: ELRA.
- Botdorf M, Riggis T (2018) When less is more: Thinner fronto-parietal cortices are associated with better forward digit span performance during early childhood. *Neuropsychologia* 121:11-18.
- Bowles AR, Chang CB, Karuzis VP (2016) Pitch Ability As an Aptitude for Tone Learning. *Language Learning* 66:774-808.
- Brodmann K (1909) Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues: Barth.
- Bruce G (1977) Swedish word accents in sentence perspective. Lund: Gleerups.
- Bruce G (1987) How floating is focal accent? In: *Nordic Prosody IV* (Gregersen K, Basbøll H, eds), pp 41-49: Odense University Press.
- Buzsáki G (2020) The Brain–Cognitive Behavior Problem: A Retrospective. *neuro* 7:ENEURO.0069-0020.2020.
- Carroll J (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 JB, Sapon S (1959) *Modern language aptitude test*. San Antonio: TX: Psychological Corporation.
- Catani M, Thiebaut de Schotten M (2008) A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex* 44:1105-1132.

- Cavanna AE, Trimble MR (2006) The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129:564-583.
- Cohen J (1977) CHAPTER 1 - The Concepts of Power Analysis. In: *Statistical Power Analysis for the Behavioral Sciences* (Cohen J, ed), pp 1-17: Academic Press.
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* 9:179-194.
- Delfino JG, Krainak DM, Flesher SA, Miller DL (2019) MRI-related FDA adverse event reports: A 10-yr review. *Medical Physics* 46:5562-5571.
- Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, 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.
- Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* 53:1-15.
- DeWitt I, Rauschecker JP (2012) Phoneme and word recognition in the auditory ventral stream. *Proc Natl Acad Sci U S A* 109:E505-514.
- DeWitt I, Rauschecker JP (2013) Wernicke's area revisited: parallel streams and word processing. *Brain Lang* 127:181-191.
- DeWitt I, Rauschecker JP (2016) Convergent evidence for the causal involvement of anterior superior temporal gyrus in auditory single-word comprehension. *Cortex* 77:164-166.
- Donders, F.C. (1969) On the speed of mental processes. *Acta Psychologica* 30: 412-31.
- Federmeier KD, Wlotko EW, De Ochoa-Dewald E, Kutas M (2007) Multiple effects of sentential constraint on word processing. *Brain Research* 1146:75-84.
- Fieremans E, Jensen JH, Helpen JA (2011) White matter characterization with diffusional kurtosis imaging. *Neuroimage* 58:177-188.
- Fischl B, Dale AM (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America* 97:11050-11055.
- Fischl B, Sereno MI, Dale AM (1999) Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System. *NeuroImage* 9:195-207.
- 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, Dale AM (2004) Automatically Parcellating the Human Cerebral Cortex. *Cerebral cortex* (New York, NY : 1991) 14:11-22.
- Friederici AD (2002) Towards a neural basis of auditory sentence processing. *Trends Cogn Sci* 6:78-84.
- Friederici AD, Gierhan SME (2013) The language network. *Current Opinion in Neurobiology* 23:250-254.
- Friederici AD, Pfeifer E, Hahne A (1993) Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research* 1:183-192.
- Gambi C, Gorrie F, Pickering MJ, Rabagliati H (2018) The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds. *Journal of experimental child psychology* 173:351-370.

- Goel V, Gold B, Kapur S, Houle S (1997) The seats of reason? An imaging study of deductive and inductive reasoning. *Neuroreport* 8:1305-1310.
- 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* 18:6-34.
- Golestani N, Paus T, Zatorre RJ (2002) Anatomical Correlates of Learning Novel Speech Sounds. *Neuron* 35:997-1010.
- Golestani N, Price CJ, Scott SK (2011) Born with an Ear for Dialects? Structural Plasticity in the Expert Phonetician Brain. *The Journal of Neuroscience* 31:4213-4220.
- Golestani N, Molko N, Dehaene S, LeBihan D, Pallier C (2007) Brain structure predicts the learning of foreign speech sounds. *Cerebral cortex (New York, NY : 1991)* 17:575-582.
- Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. *Trends Neurosci* 15:20-25.
- Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, Ogar JM, Rohrer JD, Black S, Boeve BF, Manes F, Dronkers NF, Vandenberghe R, Rascovsky K, Patterson K, Miller BL, Knopman DS, Hodges JR, Mesulam MM, Grossman M (2011) Classification of primary progressive aphasia and its variants. *Neurology* 76:1006-1014.
- Goucha T, Friederici AD (2015) The language skeleton after dissecting meaning: A functional segregation within Broca's Area. *NeuroImage* 114:294-302.
- Granena G (2013) Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Tests. In: Sensitive periods, language aptitude, and ultimate L2 attainment: John Benjamins.
- Granena G, Long MH (2013) Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research* 29:311-343.
- Graves RE (1997) The Legacy of the Wernicke-Lichtheim Model. *Journal of the History of the Neurosciences* 6:3-20.
- Hagoort P (2005) On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences* 9:416-423.
- Hahne A, Friederici AD (2002) Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research* 13:339-356.
- Haier RJ, Jung RE, Yeo RA, Head K, Alkire MT (2004) Structural brain variation and general intelligence. *NeuroImage* 23:425-433.
- 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, 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.
- Hartwig V, Giovannetti G, Vanello N, Lombardi M, Landini L, Simi S (2009) Biological effects and safety in magnetic resonance imaging: a review. *Int J Environ Res Public Health* 6:1778-1798.
- Hartwigsen G (2015) The neurophysiology of language: Insights from non-invasive brain stimulation in the healthy human brain. *Brain and Language* 148:81-94.

- Hed A, Schremm A, Horne M, Roll M (2019) Neural correlates of second language acquisition of tone-grammar associations. *The Mental Lexicon* 14:98-123.
- Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? *The Behavioral and brain sciences* 33:61-83; discussion 83-135.
- Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92:67-99.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nature Reviews Neuroscience* 8:393.
- Hirose Y, Mazuka R (2017) Exploiting Pitch Accent Information in Compound Processing: A Comparison between Adults and 6- to 7-Year-Old Children. *Language Learning and Development* 13:375-394.
- Hui ES, Cheung MM, Qi L, Wu EX (2008) Towards better MR characterization of neural tissues using directional diffusion kurtosis analysis. *NeuroImage* 42:122-134.
- Hyde KL, Zatorre RJ, Griffiths TD, Lerch JP, Peretz I (2006) Morphometry of the amusic brain: a two-site study. *Brain* 129:2562-2570.
- Hyde KL, Lerch JP, Zatorre RJ, Griffiths TD, Evans AC, Peretz I (2007) Cortical Thickness in Congenital Amusia: When Less Is Better Than More. *Journal of Neuroscience* 27:13028-13032.
- Jbabdi S, Johansen-Berg H (2011) Tractography: Where Do We Go from Here? *Brain Connectivity* 1:169-183.
- Jensen JH, Helpert JA, Ramani A, Lu H, Kaczynski K (2005) Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine* 53:1432-1440.
- Jin Y, Shi Y, Zhan L, Li J, de Zubicaray GI, McMahon KL, Martin NG, Wright MJ, Thompson PM (2012) Automatic Population HARDI White Matter Tract Clustering by Label Fusion of Multiple Tract Atlases. In: *Multimodal Brain Image Analysis* (Yap P-T, Liu T, Shen D, Westin C-F, Shen L, eds), pp 147-156. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Jin Y, Shi Y, Zhan L, Gutman BA, de Zubicaray GI, McMahon KL, Wright MJ, Toga AW, Thompson PM (2014) Automatic clustering of white matter fibers in brain diffusion MRI with an application to genetics. *NeuroImage* 100:75-90.
- Johnsrude IS, Penhune VB, Zatorre RJ (2000) Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123:155-163.
- Jung T-P, Makeig S, Humphries C, Lee T-W, McKeown MJ, Iragui V, Sejnowski TJ (2000) Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37:163-178.
- Kahn I, Andrews-Hanna JR, Vincent JL, Snyder AZ, Buckner RL (2008) Distinct cortical anatomy linked to subregions of the medial temporal lobe revealed by intrinsic functional connectivity. *Journal of neurophysiology* 100:129-139.
- Karama S, Ad-Dab'bagh Y, Haier RJ, Deary IJ, Lyttelton OC, Lepage C, Evans AC, Brain Development Cooperative G (2009) Positive association between cognitive ability and cortical thickness in a representative US sample of healthy 6 to 18 year-olds. *Intelligence* 37:145-155.
- Klein D, Mok K, Chen J-K, Watkins KE (2014) Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language* 131:20-24.

- Korb JP, Bryant RG (2002) Magnetic field dependence of proton spin-lattice relaxation times. *Magnetic resonance in medicine* 48:21-26.
- Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. *Nature Reviews Neuroscience* 11:599-605.
- Kutas M, Hillyard S (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207:203-205.
- Kutas M, Federmeier KD (2010) Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology* 62:621-647.
- Kutas M, Van Petten CK, Kluender R (2006) Chapter 17 - Psycholinguistics Electrified II (1994–2005). In: *Handbook of Psycholinguistics (Second Edition)* (Traxler MJ, Gernsbacher MA, eds), pp 659-724. London: Academic Press.
- Kyllonen PC, Zu J (2016) Use of Response Time for Measuring Cognitive Ability. *Journal of Intelligence* 4:14.
- Le Gac D (2003) Tonal alternations in Somali. In: *Fifth Conference on Afroasiatic Languages* (Lecarme J, ed), pp 287-304. Paris: John Benjamins Publishing Company.
- Lee L, Lin S (2015) The Impact of Music Activities on Foreign Language, English Learning for Young Children. In.
- Lee RR-W, Hsu C-H, Lin S-K, Wu DH, Tzeng OJ-L (2017) Learning transforms functional organization for Mandarin lexical tone discrimination in the brain: Evidence from a MEG experiment on second language learning. *Journal of Neurolinguistics* 42:124-139.
- Li S (2014) The Associations Between Language Aptitude and Second Language Grammar Acquisition: A Meta-Analytic Review of Five Decades of Research. *Applied Linguistics* 36:385-408.
- Li P, Legault J, Litcofsky KA (2014) Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex* 58:301-324.
- López-Barroso D, de Diego-Balaguer R (2017) Language Learning Variability within the Dorsal and Ventral Streams as a Cue for Compensatory Mechanisms in Aphasia Recovery. *Frontiers in Human Neuroscience* 11.
- Luck SJ (2005) *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Luck SJ (2012) Event-related potentials. In: *APA handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics.*, pp 523-546. Washington, DC, US: American Psychological Association.
- Maier-Hein KH et al. (2017) The challenge of mapping the human connectome based on diffusion tractography. *Nat Commun* 8:1349-1349.
- Malpas CB, Genc S, Saling MM, Velakoulis D, Desmond PM, O'Brien TJ (2016) MRI correlates of general intelligence in neurotypical adults. *Journal of Clinical Neuroscience* 24:128-134.
- Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele P-F, Gruetter R (2010) MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage* 49:1271-1281.
- Marslen-Wilson WD, Tyler LK (2007) Morphology, language and the brain: the decompositional substrate for language comprehension. *Philos Trans R Soc Lond B Biol Sci* 362:823-836.

- Mathys C, Loui P, Zheng X, Schlaug G (2010) Non-invasive brain stimulation applied to Heschl's gyrus modulates pitch discrimination. *Frontiers in psychology* 1:193-193.
- McGettigan C, Warren JE, Eisner F, Marshall CR, Shanmugalingam P, Scott SK (2010) Neural Correlates of Sublexical Processing in Phonological Working Memory. *Journal of Cognitive Neuroscience* 23:961-977.
- Meara P (2005) LLAMA Language Aptitude Tests: The Manual. In. University of Wales Swansea: _lognostics.
- Meara P, Milton J, Lorenzo-Dus N (2001) Language aptitude tests. Express Publishing.
- Menary K, Collins PF, Porter JN, Muetzel R, Olson EA, Kumar V, Steinbach M, Lim KO, Luciana M (2013) Associations between cortical thickness and general intelligence in children, adolescents and young adults. *Intelligence* 41:597-606.
- 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* 24:2541-2552.
- Moser E, Stahlberg F, Ladd ME, Trattnig S (2012) 7-T MR—from research to clinical applications? *NMR in Biomedicine* 25:695-716.
- 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:e89642.
- Mårtensson J, Lövdén M (2011) Do Intensive Studies of a Foreign Language Improve Associative Memory Performance? *Frontiers in Psychology* 2.
- Mårtensson J, Eriksson J, Bodammer NC, Lindgren M, Johansson M, Nyberg L, Lövdén M (2012) Growth of language-related brain areas after foreign language learning. *NeuroImage* 63:240-244.
- 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.
- Natu VS, Gomez J, Barnett M, Jeska B, Kirilina E, Jaeger C, Zhen Z, Cox S, Weiner KS, Weiskopf N, Grill-Spector K (2019) Apparent thinning of human visual cortex during childhood is associated with myelination. *Proceedings of the National Academy of Sciences* 116:20750-20759.
- Neville H, Nicol JL, Barss A, Forster KI, Garrett MF (1991) Syntactically based sentence processing classes: evidence from event-related brain potentials. *J Cogn Neurosci* 3:151-165.
- Novén M, Schremm A, Horne M, Roll M (2021) Cortical thickness and surface area of left anterior temporal areas affects processing of phonological cues to morphosyntax. *Brain Research* 1750:147150.
- 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.
- O'Donnell LJ, Westin C (2007) Automatic Tractography Segmentation Using a High-Dimensional White Matter Atlas. *IEEE Transactions on Medical Imaging* 26:1562-1575.
- Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America* 87:9868-9872.

- Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences* 89:5951-5955.
- Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9:97-113.
- Osterhout L, Holcomb PJ (1992) Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31:785-806.
- Panizzon MS, Fennema-Notestine C, Eyler LT, Jernigan TL, Prom-Wormley E, Neale M, Jacobson K, Lyons MJ, Grant MD, Franz CE, Xian H, Tsuang M, Fischl B, Seidman L, Dale A, Kremen WS (2009) Distinct Genetic Influences on Cortical Surface Area and Cortical Thickness. *Cerebral Cortex* 19:2728-2735.
- Panych LP, Madore B (2018) The physics of MRI safety. *Journal of Magnetic Resonance Imaging* 47:28-43.
- Penke M, Weyerts H, Gross M, Zander E, Munte TF, Clahsen H (1997) How the brain processes complex words: an event-related potential study of German verb inflections. *Brain research Cognitive brain research* 6:37-52.
- Peretz I, Ayotte J, Zatorre RJ, Mehler J, Ahad P, Penhune VB, Jutras Bt (2002) Congenital Amusia: A Disorder of Fine-Grained Pitch Discrimination. *Neuron* 33:185-191.
- Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Di Chiro G (1996) Diffusion tensor MR imaging of the human brain. *Radiology* 201:637-648.
- Pinker S (1999) *Words and rules: The ingredients of language*. London: Weidenfeld & Nicolson.
- Poeppl D (2014) The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology* 28:142-149.
- Poeppl D, Adolfs F (2020) Against the Epistemological Primacy of the Hardware: The Brain from Inside Out, Turned Upside Down. *neuro* 7:ENEURO.0215-0220.2020.
- Posedel J, Emery L, Souza B, Fountain C (2011) Pitch perception, working memory, and second-language phonological production. *Psychology of Music* 40:508-517.
- Pujol J, Deus J, Losilla JM, Capdevila A (1999) Cerebral lateralization of language in normal left-handed people studied by functional MRI. *Neurology* 52:1038-1038.
- Pulvermüller F, Shtyrov Y (2003) Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *Neuroimage* 20:159-172.
- Purves D, Augustine GJ, Fitzpatrick D, Hall WC, LaMantia A-S, McNamara JO, Williams SM (2004) *Neuroscience*, 5th ed. Sunderland, MA, US: Sinauer Associates.
- Qi Z, Han M, Garel K, San Chen E, Gabrieli JDE (2015) White-matter structure in the right hemisphere predicts Mandarin Chinese learning success. *Journal of Neurolinguistics* 33:14-28.
- Rakic P (1988) Specification of cerebral cortical areas. *Science* 241:170-176.
- Rakic P (2000) Radial Unit Hypothesis of Neocortical Expansion. In: *Evolutionary Developmental Biology of the Cerebral Cortex* (Bock GR, Cardew G, eds), pp 30-45: John Wiley & Sons Ltd.
- Raven J (2000) The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology* 41:1-48.
- Reber AS (1967) Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6:855-863.

- Redpath TW (1998) Signal-to-noise ratio in MRI. *The British Journal of Radiology* 71:704-707.
- Reuter M, Schmansky NJ, Rosas HD, Fischl B (2012) Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* 61:1402-1418.
- Riad T (1998) The Origin of Scandinavian Tone Accents. *Diachronica* 15:63-98.
- Riad T (2012) Culminativity, stress and tone accent in Central Swedish. *Lingua* 122:1352-1379.
- Riad T (2014) *The phonology of Swedish*: Oxford University Press.
- Rischel J (1963) Morphemic Tone and Word Tone in Eastern Norwegian. *Phonetica* 10:154-164.
- Roberts DC, Marcelli V, Gillen Joseph S, Carey John P, Della Santina Charles C, Zee David S (2011) MRI Magnetic Field Stimulates Rotational Sensors of the Brain. *Current Biology* 21:1635-1640.
- Robinson P (2010) Implicit Artificial Grammar and Incidental Natural Second Language Learning: How Comparable Are They? *Language Learning* 60:245-263.
- Rodriguez SM, Archila-Suerte P, Vaughn KA, Chiarello C, Hernandez AE (2018) Anterior insular thickness predicts speech sound learning ability in bilinguals. *NeuroImage* 165:278-284.
- Rogers V, Meara P, Barnett-Legh T, Curry C, Davie E (2017) Examining the LLAMA aptitude tests. *Journal of the European Second Language Association* 1:49-60.
- Rogers VE, Meara P, Aspinall R, Fallon L, Goss T, Key E, Thomas R (2016) Testing aptitude: Investigating Meara's (2005) LLAMA tests. *EUROSLA Yearbook* 16:179-210.
- Roll M (2015) A neurolinguistic study of South Swedish word accents: Electrical brain potentials in nouns and verbs. *Nordic Journal of Linguistics* 38:149-162.
- Roll M, Horne M, Lindgren M (2010) Word accents and morphology—ERPs of Swedish word processing. *Brain Research* 1330:114-123.
- Roll M, Söderström P, Horne M (2013) Word-stem tones cue suffixes in the brain. *Brain Research* 1520:116-120.
- Roll M, Söderström P, Frid J, Mannfolk P, Horne M (2017) Forehearing words: Pre-activation of word endings at word onset. *Neuroscience Letters* 658:57-61.
- Roll M, Söderström P, Mannfolk P, Shtyrov Y, Johansson M, van Westen D, Horne M (2015) Word tones cueing morphosyntactic structure: Neuroanatomical substrates and activation time-course assessed by EEG and fMRI. *Brain and Language* 150:14-21.
- Rolle NR (2018) Grammatical tone: Typology and theory. In: *Linguistics*. Berkely: University of California.
- Rooney WD, Johnson G, Li X, Cohen ER, Kim S-G, Ugurbil K, Springer Jr. CS (2007) Magnetic field and tissue dependencies of human brain longitudinal $^1\text{H}_2\text{O}$ relaxation in vivo. *Magnetic Resonance in Medicine* 57:308-318.
- Rosas HD, Liu AK, Hersch S, Glessner M, Ferrante RJ, Salat DH, van der Kouwe A, Jenkins BG, Dale AM, Fischl B (2002) Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology* 58:695-701.
- Sadeghi-Tarakameh A, DelaBarre L, Lagore RL, Torrado-Carvajal A, Wu X, Grant A, Adriany G, Metzger GJ, Van de Moortele P-F, Ugurbil K, Atalar E, Eryaman Y (2020) In vivo human head MRI at 10.5T: A radiofrequency safety study and preliminary imaging results. *Magnetic Resonance in Medicine* 84:484-496.

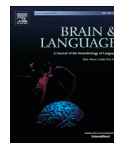
- Saur D, Kreher BW, Schnell S, Kümmerer D, Kellmeyer P, Vry M-S, Umarova R, Musso M, Glauche V, Abel S, Huber W, Rijntjes M, Hennig J, Weiller C (2008) Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America* 105:18035-18040.
- Schnack HG, van Haren NEM, Brouwer RM, Evans A, Durston S, Boomsma DI, Kahn RS, Hulshoff Pol HE (2015) Changes in Thickness and Surface Area of the Human Cortex and Their Relationship with Intelligence. *Cerebral Cortex* 25:1608-1617.
- Schnikpe DL, Scrams DJ (1999) Exploring Issues of Test Taker Behavior: Insights Gained from Response-Time Analyses. Law School Admission Council Computerized Testing Report. LSAC Research Report Series. In.
- Schremm A, Novén M, Horne M, Söderström P, van Westen D, Roll M (2018) Cortical thickness of planum temporale and pars opercularis in native language tone processing. *Brain and Language* 176:42-47.
- Schwartz MF, Faseyitan O, Kim J, Coslett HB (2012) The dorsal stream contribution to phonological retrieval in object naming. *Brain* 135:3799-3814.
- Scott TL, Perrachione TK (2019) Common cortical architectures for phonological working memory identified in individual brains. *NeuroImage* 202:116096.
- Sebastián-Gallés N, Soriano-Mas C, Baus C, Díaz B, Ressel V, Pallier C, Costa A, Pujol J (2012) Neuroanatomical markers of individual differences in native and non-native vowel perception. *Journal of Neurolinguistics* 25:150-162.
- Slevc LR, Miyake A (2006) Individual Differences in Second-Language Proficiency. *Psychological Science* 17:675-681.
- Sowell ER, Thompson PM, Leonard CM, Welcome SE, Kan E, Toga AW (2004) Longitudinal Mapping of Cortical Thickness and Brain Growth in Normal Children. *The Journal of Neuroscience* 24:8223.
- Sporns O (2011) The human connectome: a complex network. *Annals of the New York Academy of Sciences* 1224:109-125.
- 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.
- Söderström P, Roll M, Horne M (2012) Processing morphologically conditioned word accents. *The Mental Lexicon* 7:77-89.
- Söderström P, Horne M, Roll M (2017a) Stem Tones Pre-activate Suffixes in the Brain. *Journal of Psycholinguistic Research* 46:271-280.
- Söderström P, Horne M, Frid J, Roll M (2016) Pre-Activation Negativity (PrAN) in Brain Potentials to Unfolding Words. *Frontiers in Human Neuroscience* 10.
- Söderström P, Horne M, Mannfolk P, van Westen D, Roll M (2017b) Tone-grammar association within words: Concurrent ERP and fMRI show rapid neural pre-activation and involvement of left inferior frontal gyrus in pseudoword processing. *Brain and Language* 174:119-126.
- Söderström P, Horne M, Mannfolk P, van Westen D, Roll M (2018) Rapid syntactic pre-activation in Broca's area: Concurrent electrophysiological and haemodynamic recordings. *Brain Research* 1697:76-82.
- Tamnes CK, Fjell AM, Østby Y, Westlye LT, Due-Tønnessen P, Bjørnerud A, Walhovd KB (2011) The brain dynamics of intellectual development: Waxing and waning white and gray matter. *Neuropsychologia* 49:3605-3611.

- Tournier J-D, Calamante F, Connelly A (2012) MRtrix: Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology* 22:53-66.
- Tournier JD, Calamante F, Connelly A (2007) Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage* 35:1459-1472.
- Tremblay P, Dick AS (2016) Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language* 162:60-71.
- Turker S, Reiterer SM, 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.
- Turker S, Reiterer SM, Schneider P, Seither-Preisler A (2019) Auditory Cortex Morphology Predicts Language Learning Potential in Children and Teenagers. *Frontiers in Neuroscience* 13.
- Uddén J, Männel C (2018) Artificial grammar learning and its neurobiology in relation to language processing and development. In: *The Oxford Handbook of Psycholinguistics*, 2 Edition (Rueschemeyer SA, Gaskell MG, eds), pp 755-783. Oxford: Oxford University Press.
- Ullman MT (2001) A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience* 2:717-726.
- Ullman MT (2004) Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92:231-270.
- Unsworth N, Heitz RP, Schrock JC, Engle RW (2005) An automated version of the operation span task. *Behavior Research Methods* 37:498-505.
- Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K (2013) The WU-Minn Human Connectome Project: An overview. *NeuroImage* 80:62-79.
- Vaughan HG (1982) The neural origins of human event-related potentials. *Annals of the New York Academy of Sciences* 388:125-138.
- Veroude K, Norris DG, Shumskaya E, Gullberg M, Indefrey P (2010) Functional connectivity between brain regions involved in learning words of a new language. *Brain and Language* 113:21-27.
- Vincent JL, Kahn I, Snyder AZ, Raichle ME, Buckner RL (2008) Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of neurophysiology* 100:3328-3342.
- Vincent JL, Snyder AZ, Fox MD, Shannon BJ, Andrews JR, Raichle ME, Buckner RL (2006) Coherent spontaneous activity identifies a hippocampal-parietal memory network. *J Neurophysiol* 96:3517-3531.
- Vuoksima E, Panizzon MS, Chen C-H, Fiecas M, Eyer LT, Fennema-Notestine C, Hagler DJ, Jr., Franz CE, Jak AJ, Lyons MJ, Neale MC, Rinker DA, Thompson WK, Tsuang MT, Dale AM, Kremen WS (2016) Is bigger always better? The importance of cortical configuration with respect to cognitive ability. *NeuroImage* 129:356-366.
- Vuoksima E, Panizzon MS, Chen C-H, Fiecas M, Eyer LT, Fennema-Notestine C, Hagler DJ, Fischl B, Franz CE, Jak A, Lyons MJ, Neale MC, Rinker DA, Thompson WK, Tsuang MT, Dale AM, Kremen WS (2015) The Genetic Association Between Neocortical Volume and General Cognitive Ability Is Driven by Global Surface Area Rather Than Thickness. *Cerebral cortex (New York, NY : 1991)* 25:2127-2137.

- Wakana S, Jiang H, Nagae-Poetscher LM, van Zijl PC, Mori S (2004) Fiber tract-based atlas of human white matter anatomy. *Radiology* 230:77-87.
- Wasserthal J, Neher PF, Maier-Hein KH (2018a) TractSeg - Fast and accurate white matter tract segmentation. *NeuroImage* 183:239-253.
- Wasserthal J, Neher PF, Maier-Hein KH (2018b) Tract Orientation Mapping for Bundle-Specific Tractography. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, eds), pp 36-44. Cham: Springer International Publishing.
- Wasserthal J, Neher PF, Hirjak D, Maier-Hein KH (2019) Combined tract segmentation and orientation mapping for bundle-specific tractography. *Medical image analysis* 58:101559.
- Wen Z, Biedroń A, Skehan P (2016) Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching* 50:1-31.
- Weyerts H, Penke M, Dohrn U, Clahsen H, Munte TF (1997) Brain potentials indicate differences between regular and irregular German plurals. *Neuroreport* 8:957-962.
- Wong FCK, Chandrasekaran B, Garibaldi K, Wong PCM (2011) White Matter Anisotropy in the Ventral Language Pathway Predicts Sound-to-Word Learning Success. *The Journal of Neuroscience* 31:8780-8785.
- Wong PCM, Perrachione TK, Parrish TB (2007a) Neural characteristics of successful and less successful speech and word learning in adults. *Hum Brain Mapp* 28:995-1006.
- Wong PCM, Skoe E, Russo NM, Dees T, Kraus N (2007b) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience* 10:420-422.
- Wong PCM, Warrier CM, Penhune VB, Roy AK, Sadehh A, Parrish TB, Zatorre RJ (2008) Volume of Left Heschl's Gyrus and Linguistic Pitch Learning. *Cerebral Cortex* 18:828-836.
- Xiang H, Dediu D, Roberts L, Oort Ev, Norris DG, Hagoort P (2012) The Structural Connectivity Underpinning Language Aptitude, Working Memory, and IQ in the Perisylvian Language Network. *Language Learning* 62:110-130.
- Zaretskaya N, Fischl B, Reuter M, Renvall V, Polimeni JR (2018) Advantages of cortical surface reconstruction using submillimeter 7 T MEMPRAGE. *Neuroimage* 165:11-26.
- Zatorre RJ, Evans A, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846-849.
- Zatorre RJ, Fields RD, Johansen-Berg H (2012) Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nat Neurosci* 15:528-536.

Study I





Short communication

Cortical thickness of planum temporale and pars opercularis in native language tone processing



Andrea Schremm^{a,*}, Mikael Novén^a, Merle Horne^a, Pelle Söderström^a, Danielle van Westen^b, Mikael Roll^a

^a Department of Linguistics and Phonetics, Lund University, Box 201, 22100 Lund, Sweden

^b Department of Diagnostic Radiology, Clinical Sciences, Lund University, Box 201, 22100 Lund, Sweden

ARTICLE INFO

Keywords:

Cortical thickness
Linguistic tone
Planum temporale
Pars opercularis

ABSTRACT

The present study investigated the relationship between linguistic tone processing and cortical thickness of bilateral planum temporale (PT) and pars opercularis of the inferior frontal gyrus (IFGpo). Swedish tones on word stems function as cues to upcoming endings. Correlating structural brain imaging data with participants' response time patterns for suffixes, we found that thicker cortex in the left PT was associated with greater reliance on tones to anticipate upcoming inflections on real words. On inflected pseudoword stems, however, the cortical thickness of left IFGpo was associated with tone-suffix processing. Thus cortical thickness of the left PT might play a role in processing tones as part of stored representations for familiar speech segments, most likely when inflected forms are accessed as whole words. In the absence of stored representations, listeners might need to rely on morphosyntactic rules specifying tone-suffix associations, potentially facilitated by greater cortical thickness of left IFGpo.

1. Introduction

General cognitive abilities have been found to be related to the cortical thickness in associated brain areas (Karama et al., 2009). Similarly, the structure of primary auditory regions in the left hemisphere has been observed to affect non-native lexical tone-learning ability (Wong et al., 2008). To date, however, it is not known how individual differences in brain morphology are related to word tone processing in native speakers. It is difficult to quantify the use of lexical tones in languages such as Chinese, since tones change word meaning like any other phonemes: as a word pronounced with a certain tone pattern is either identified as a specific lexical item or not, variation in tone processing performance beyond word recognition success is not straightforward to observe. In Swedish and Norwegian, however, tones on word stems are associated with suffixes (Riad, 2014). For example, the word *bil* 'car' is pronounced with accent 1 (a low tone) if it appears with the singular suffix *-en*, but with accent 2 (a high tone) if it ends in the plural suffix *-ar*. Accordingly, tones can be used to speed up suffix processing, generating an increased pre-activation negativity (PrAN) in the electrical brain potentials (Roll et al., 2015; Söderström, Horne, Mannfolk, van Westen, & Roll, 2017). PrAN has been found to increase gradually as the number of possible word completions decreased,

suggesting that it reflects predictive activation of memory traces, modulated by the certainty that an upcoming continuation will appear (Söderström, Horne, Frid, & Roll, 2016). Swedish tones are thus not associated with a categorical lexical distinction but with a graded process related to the degree of pre-activation, providing a unique opportunity to quantify the use of tones in native speech comprehension. In an experimental context, if the wrong tone-suffix association is presented, participants would be expected to take a longer time to recognize the suffix the more they rely on the tonal cue to process the ending. Thus, macrostructural differences in auditory association areas in the left planum temporale (PT) region, previously observed to be involved in processing tones on inflected Swedish word stems (Roll et al., 2015), as well as in Chinese and Thai tone processing (Xu et al., 2006), might be expected to be related to performance. Since Swedish speakers have been reported to process most inflected real words similar to monomorphemic ones, i.e. by accessing full forms in the mental lexicon (Lehtonen, Niska, Wande, Niemi, & Laine, 2006), reliance on the tonal cue might depend on the degree to which the tone is incorporated in a whole word representation. In the absence of lexical content, however, participants can be assumed to rely on morphosyntactic rules for stem tone-suffix combinations in order to optimally process inflections, since whole-word representations are unavailable.

* Corresponding author at: Department of Linguistics and Phonetics, Centre for Languages and Literature, Lund University, Box 201, 221 00 Lund, Sweden.

E-mail addresses: andrea.schremm@ling.lu.se (A. Schremm), mikael.noven@ling.lu.se (M. Novén), merle.horne@ling.lu.se (M. Horne), pelle.soderstrom@ling.lu.se (P. Söderström), danielle.van_westen@med.lu.se (D. van Westen), mikael.roll@ling.lu.se (M. Roll).

<https://doi.org/10.1016/j.bandl.2017.12.001>

Received 13 April 2017; Received in revised form 13 October 2017; Accepted 1 December 2017

Available online 07 December 2017

0093-934X/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

That Swedish tones also activate areas associated with the processing of word structure was particularly obvious in a study involving pseudowords, where haemodynamic responses indicated a neural network partially distinct from that observed for real words (Roll et al., 2015), with activation centering around the pars opercularis of the left inferior frontal gyrus (IFGpo) (Söderström et al., 2017). The present study tested the relation between cortical thickness in bilateral PT and IFGpo and native speakers' use of tones for suffix pre-activation in both real words and pseudowords.

Measures of cortical thickness can be obtained by calculating the shortest distance between the white matter surface to the pial surface, constituting the border between the grey matter and the cerebrospinal fluid (Fischl & Dale, 2000). As postulated by the radial unit hypothesis (Rakic, 1988), the human cortex is characterized by a columnar organization, established during embryonic development by migration of cells along radial glial guides towards their final location. Cortical thickness can thus be related to the number and size of neurons within a column (Rakic, 1988). The number and spacing of columns affect cortical surface area instead, implicating area as a distinct feature of cortical structure (Meyer, Liem, Hirsiger, Jäncke, & Hänggi, 2014; Rakic, 1988).

Structural brain imaging data and response time values collected in two previous experiments with two different participant groups, focusing on processing Swedish tone-suffix associations in real words (Roll et al., 2015) as opposed to pseudowords (Söderström et al., 2017), were analysed. In both experiments, test items involved stimuli carrying a singular or plural suffix, which was either validly cued by its preceding associated tone or invalidly cued by a tone associated with a different suffix. A relative increase in response time from validly to invalidly cued suffixes can be assumed to reflect the degree to which the perceived ending disconfirmed the tone-induced prediction: the more the listener anticipated the validly cued number inflection based on the stem tone, the greater the expected processing cost for the invalidly cued suffix. In the present study, we investigated the association between response times and cortical thickness of bilateral PT and IFGpo. It was hypothesized that greater reliance on tonal cues on real words, when tonal patterns are incorporated into whole word representations, would be associated with variation in cortical thickness in left PT, in accordance with the central role of this area in tone processing in Swedish words (Roll et al., 2015). Pseudowords, however, do not have any existing word representations, and therefore their processing cannot rely on a tone-word form association. Since the experimental task involved a decision between singular versus plural meanings carried by inflectional suffixes but did not require interpretation of the meaning of target word stems, response times are not assumed to primarily reflect different semantic processing of real words and pseudowords due to the presence versus absence of familiar semantic content. Instead, pseudoword response time patterns would rather depend on the efficiency of tone-induced morphosyntactic rule application. Decompositional morphosyntactic processes underlying the extraction and interpretation of regular inflectional affixes seem to rely on an intact left inferior frontal gyrus (IFG) (Bozic, Fonteneau, Su, & Marslen-Wilson, 2015). Generally, the left IFG has been argued to function as a unification space, maintaining fragments of syntactic, semantic and phonological information retrieved from memory and assembling these into coherent representations. Unification processes are assumed to take place even below the phrasal level, by which word forms are composed from and decomposed into stem and affix parts (Hagoort, 2013). Previous studies indeed found areas of left IFG, specifically the pars opercularis or pars triangularis, to be crucially involved in inflectional morpheme processing (Bozic et al., 2015; Tyler, Stamatakis, Post, Randall, & Marslen-Wilson, 2005). The present study focuses on the IFGpo, the area that showed strong activation for processing inflected Swedish pseudowords (Söderström et al., 2017).

In order to develop an expectation for the direction of the relationship between tone processing and cortical structure,

electrophysiological measurements obtained in Roll et al. (2015) as well as Söderström et al. (2017) were related to cortical thickness values (Supplementary material). As mentioned above, greater tone-induced suffix pre-activation was associated with an increased negativity in the electrical brain response in these studies. Electrophysiological measurements rely on synchronous firing of neurons; thus, increased electrophysiological signal might conceivably be generated by a larger number of neurons within a thicker cortex, or alternatively by more synchronized activity of neurons in a fine-tuned network of a thinner cortex (Liem, Zaehle, Burkhard, Jäncke, & Meyer, 2012). The present analysis indicated that an increased pre-activation negativity for real words tended to be associated with thicker cortex in the left PT. We therefore predicted a positive association between tone processing performance and cortical thickness in the present study. As a complementary measurement to cortical thickness, surface area of bilateral PT and IFGpo was also related to tone processing performance in a supplementary analysis (see Supplementary material).

2. Results

2.1. Real word processing

Individual participants' RT advantage for valid over invalid suffixes showed a significant positive correlation with average cortical thickness in the left PT ($r = .559, p = .030$), indicating that the thicker the cortex in left PT, the greater was the RT increase for invalidly cued suffixes (see Fig. 1). Cortical thickness in the left IFGpo did not correlate with RTs ($r = .134, p = .634$). No significant correlation was obtained between RTs and right PT ($r = .197, p = .481$) or right IFGpo ($r = -.097, p = .732$).

Participants responded significantly faster ($t(14) = 7.348, p < .001$) to validly ($M = 584$ ms, $SD = 172$) as compared to invalidly ($M = 656$ ms, $SD = 174$) cued suffixes. As a follow-up analysis, participants were divided into two groups of equal size ($n = 7$) based on their mean cortical thickness in the left PT, resulting in a relatively thinner (2.117–2.465 mm) and a relatively thicker cortex group (2.590–2.912 mm). An independent-samples *t*-test conducted on average RT speed for validly cued suffixes indicated marginally faster RTs ($t(12) = 1.813, p = .054$) for the thicker PT cortex group ($M = 504$ ms, $SD = 89$) as compared to the thinner PT cortex group ($M = 664$ ms, $SD = 215$).

2.2. Pseudoword processing

There was a significant positive correlation between RT advantage for validly versus invalidly cued suffixes and average cortical thickness of the left IFGpo ($r = .492, p = .045$) (see Fig. 2). No significant correlation was obtained between RTs and left PT ($r = .071, p = .787$). There was no significant correlation between RTs and cortical thickness in right PT ($r = .135, p = .606$) or right IFGpo ($r = .303, p = .237$).

RTs were significantly faster ($t(16) = 6.497, p < .001$) to validly cued suffixes ($M = 888$ ms, $SD = 258$) than to invalidly cued suffixes ($M = 979$ ms, $SD = 232$). Subsequently, a relatively thinner cortex group (2.426–2.742 mm, $n = 8$) and a relatively thicker cortex group (2.758–2.992 mm, $n = 8$) were created based on participants' average cortical thickness in the left IFGpo. Results of an independent-samples *t*-test showed significantly faster RTs to validly cued suffixes ($t(14) = 1.926, p = .039$) in the thicker cortex group ($M = 763$ ms, $SD = 288$) relative to the thinner cortex group ($M = 997$ ms, $SD = 188$).

3. Discussion

Results indicated a relationship between linguistic tone processing in native speakers and cortical thickness of specific brain areas assumed to subservise these processes. Tones in Swedish are realized on word

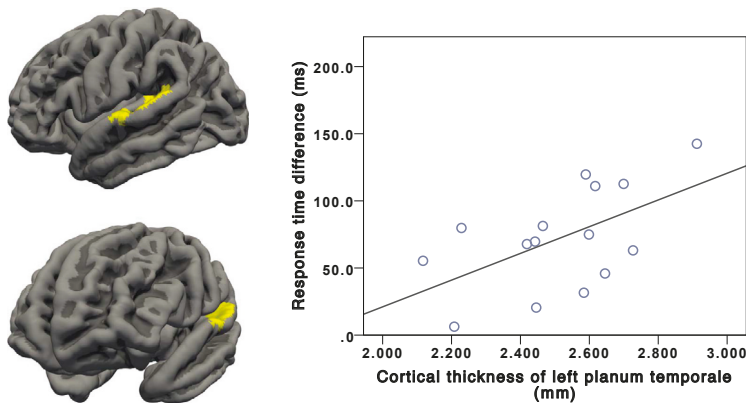


Fig. 1. Cortical thickness of left planum temporale and tone processing in real words. Left planum temporale presented on the pial surface of Freesurfer common space object fsaverage (left). Correlation between cortical thickness of left planum temporale and response time difference between suffixes invalidly vs. validly cued by the preceding tone ($r = .559$), indicating that the greater the response time increase was for invalidly cued suffixes on real word stems, the thicker the cortex was in the left planum temporale (right).

stems, and like lexical tones, are modulated by PT in the left hemisphere (Moen, 1993; Roll et al., 2015). Swedish tones function as cues to upcoming suffixes, and thus they engage left frontal cortical areas implicated in regular inflectional morpheme processing as well (Bozic et al., 2015; Tyler et al., 2005), with left IFGp emerging as the major site of activation when tone-suffix connections were implemented on meaningless pseudowords (Söderström et al., 2017). In the present study, the degree to which listeners relied on tones to anticipate morphosyntactic structure was quantified by measuring the relative response time increase for invalid tone-suffix associations. Results showed that the thicker the cortex was in the left PT, the greater the RT increase was for invalidly cued suffixes on real words. No such correlation was found for inflected pseudowords, where larger RT increase for invalidly cued suffixes was instead associated with greater cortical thickness of left IFGp.

These findings suggest that relatively greater cortical thickness of left PT is related to tonal cue processing, in ways specific to familiar lexical items. Furthermore, the same area showed functional activation for tone processing in the same subjects (Roll et al., 2015). Indeed, regions of the posterior superior temporal gyrus (STG) have been implicated in higher-level processes of acoustic analysis during speech perception (Chang et al., 2010; Xu et al., 2006) and left PT has been found to respond to phonemic changes in the native language

(Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003). The present findings indicate that the nature of the representations involved in processing related to the cortical thickness of left PT might correspond to linguistic units larger than phonemes, since the pseudowords tested consisted of Swedish phonemes, just like real word stimuli. For instance, one might assume that listeners develop memory traces for frequently occurring patterns such as syllables, supporting rapid analysis of native language input. This would constitute an important difference between the target words of the two experiments as none of the pseudoword stem syllables used in the stimulus material appeared in any real Swedish words. Thus, cortical thickness of left PT might play a role in decoding the speech signal in terms of chunks larger than phonemes – possibly syllables – incorporating tonal information in Swedish, when linguistically relevant tone patterns are processed as part of a left-lateralized network. From this perspective, the cortical thickness of left PT might be related to the efficiency with which tonal information is accessed and, in turn, tone-associated endings are subsequently pre-activated on real words, by supporting recognition of familiar sound patterns involving tones. In line with this assumption, cortical thickness of left PT was associated not only with the degree to which listeners used tonal information to activate real words but also, for the subgroup of participants with relatively thicker left PT cortex, showed a tendency to be related to faster suffix recognition as

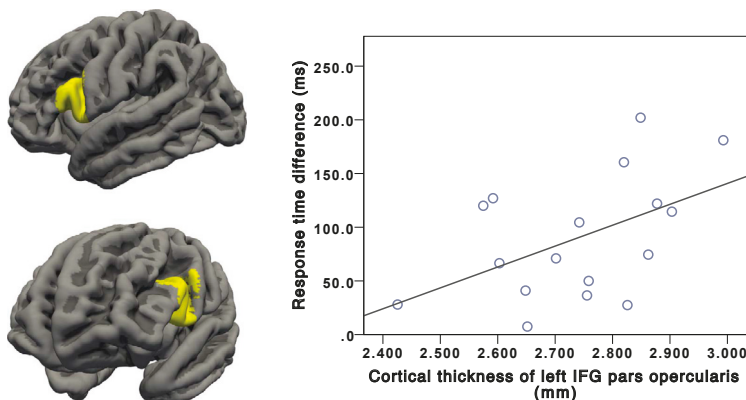


Fig. 2. Cortical thickness of left inferior frontal gyrus (IFG) pars opercularis and tone processing in pseudowords. Left IFG pars opercularis presented on the pial surface of Freesurfer common space object fsaverage (left). There was a positive correlation between cortical thickness of left IFG pars opercularis and response time advantage for validly over invalidly cued suffixes on pseudoword stems ($r = .492$) (right).

compared to the subgroup with thinner left PT cortex.

Swedish word accents are realized over several segments, in terms of relative alternations in fundamental frequency, constituting slowly changing acoustic features as compared to consonant transitions. According to the AST hypothesis (Poeppel, 2003), the duration of the integration windows used to chunk the incoming speech signal in non-primary auditory areas of the left hemisphere is around 20–50 ms, underlying sensitivity to fast changing segmental information, such as consonant transitions. Slower, 150–250 ms time windows in the right hemisphere efficiently capture suprasegmental information in the speech signal, including tones and intonation. In line with this proposal, gradually decreasing the integrity of slowly changing acoustic information in the speech input was reported to shift lateralization in the PT to the right hemisphere (Liem, Hürschler, Jäncke, & Meyer, 2014).

Harasty, Seldon, Chan, Halliday, and Harding (2003) suggested that an expansion of the cortex in the left PT, possibly driven by an increase in the underlying white matter volume, could account for the well-established leftward macrostructural asymmetry of the PT. The expansion resulted in a larger surface area but thinner cortex on the left relative to smaller surface area but thicker cortex on the right (Harasty et al., 2003). Larger surface area of the left auditory-related region has been associated with the left-hemispheric advantage for rapid acoustic analysis: in the expanded cortex, the distance between neuronal columns might be greater and dendritic trees less overlapping. In such an organization, columns might be able to function more independently, performing finely differentiated processing of the incoming auditory signal (Harasty et al., 2003; Meyer et al., 2014). Conversely, greater overlap among columns in the relatively thicker right hemisphere could result in more holistic and temporally coarse-grained analysis (Harasty et al., 2003).

Thus, auditory-related regions of the right hemisphere might be fine-tuned for processing the kind of spectral information that differentiates Swedish word accents. Nevertheless, due to their strong integration in a left-lateralized morphosyntactic system, word accents appear to be processed predominantly on the left side (Roll et al., 2015). The AST hypothesis conceptualizes hemispheric differences regarding temporal resolution as a relative phenomenon, since both hemispheres are assumed to contain neuronal populations underlying different, shorter or longer, integration windows (Poeppel, 2003). Therefore, one might speculate that the prevalence of neural organization in the left PT tuned to slower acoustic cues – otherwise more typical of the right hemisphere homologue – is somewhat greater in those with thicker left PT cortex in the present study, resulting in more efficient left-hemispheric processing of tonal information, and, in turn, facilitated access to word forms with endings cued by the stem tone. Previous results reporting a positive association between cortical thickness of the right auditory areas and relative pitch task performance (Foster & Zatorre, 2010) indicate that structural variation associated with the ability to analyse slower acoustic cues might be picked up by cortical thickness measurements.

Although pseudoword stems tested in the present study could not be processed in terms of familiar syllables, tones realized on such stems still pre-activated suffixes. This was reflected in an increase in RTs for invalidly cued endings relative to validly cued continuations. Moreover, tones on pseudowords were observed to generate an increased pre-activation negativity (PrAN) in the electrical brain signal (Söderström et al., 2017). The correlation found here between tone processing and cortical thickness of left IFGpo, a region associated with morphosyntactic analysis among other functions (Bozic et al., 2015; Tyler et al., 2005), might indicate the involvement of a morphosyntactic rule underlying the pre-activation process. Haemodynamic responses of the same participants also indicated strong functional activation of left IFGpo during pseudoword processing (Söderström et al., 2017). From this perspective, cortical thickness of left IFGpo could be assumed to play a role in facilitated pre-activation of suffixes, through efficient application of the morphosyntactic rule specifying the relevant tone-

suffix associations, subsequently speeding up the processing of an expected ending in the input. Also, strong pre-activation of the anticipated inflection might even support rapid morphological decomposition of pseudowords into stem and suffix parts, by providing reliable cues to stem boundaries in items for which no stored lexical representations exist. Such processes can be assumed to be essential for interpreting the pseudoword forms tested, since the task crucially depended on the ability to extract the number inflection from the otherwise meaningless string of segments. The subgroup with thicker cortex in left IFGpo was indeed generally faster at responding to validly cued suffixes than the thinner cortex group.

As is the case with PT, leftward asymmetry of IFGpo has been reported; however, its relation to functional language lateralization is far from clear (Keller, Crow, Foundas, Amunts, & Roberts, 2009), and it is difficult to speculate how thicker cortex in this area might support inflected pseudoword processing. Based on the discussion above, facilitated responses to word endings were presumably based on an abstracted connection between tone patterns and suffixes, implicating some form of higher level processing of the input. This might be related to previous results showing a positive connection between cortical thickness in association areas and performance of higher-level cognitive skills assumed to underlie general intelligence (Karama et al., 2009).

No significant relation was found between cortical thickness in the left PT and tone processing on pseudowords. Nevertheless, smaller cortical surface area in the left PT was associated with greater response time increase from validly to invalidly cued suffixes on pseudowords, suggesting increased reliance on tonal cues to anticipate upcoming endings (Supplementary material). This was the only significant correlation obtained with surface area in the present study, which might be assumed to reflect in part the contribution of left PT to the prelexical processing of linguistically relevant tones (Xu et al., 2006) since effective recognition and discrimination of tonal patterns necessarily precedes activation of the cued morphosyntactic information. This result is in line with the suggestion that a neural organization associated with a relatively smaller surface area, and in turn with greater overlap between cortical columns, might be especially adapt at decoding suprasegmental information (Harasty et al., 2003). The fact that no significant relationship was observed with surface area in the real word experiment could be taken to indicate certain differences in the specific cortical features of the left PT that are associated with facilitated tone processing on real words versus pseudowords. Based on the discussion above, this difference could be tentatively related to the presence versus absence of stored memory representations for syllables or whole words, which include the stem tone.

The fact that no correlation was found between cortical thickness of left IFGpo and tone processing on real words could be taken to indicate that morphosyntactic regularities and compositional processing might play a smaller role in suffix pre-activation on familiar word stems. One possible explanation is that more than one route is available for the predictive processing of tones on inflected words. This would be in line with dual-system models distinguishing between two mechanisms for the production and comprehension of morphologically complex items (Pinker & Ullman, 2002): through decomposition into stem and affix parts based on the application of productive morphosyntactic rules, which combine morphemes into complex words, or by direct access to whole-word representations of inflected forms in the mental lexicon. The decompositional route is argued to be implemented in a fronto-striatal brain network whereas whole-word access has been associated with a largely temporal system (Ullman, 2004), which is consistent with the frontal versus temporal distinction found in the present study.

Since different individuals were tested in the real word and the pseudoword experiment, one cannot exclude the possibility that the present results might in fact to some degree reflect anatomical variations between the groups. Also, these findings do not enable us to determine the source of the observed variation in cortical thickness across

participants, which might reflect genetic, environmental and/or experiential influences (e.g. Chiarello, Vazquez, Felton, & McDowell, 2016; Panizzon et al., 2009), including adaption to the demands of processing linguistic tones in the native language. Longitudinal studies involving focused training in word accent processing and tracking associated changes in cortical thickness and surface area could, therefore, further clarify the role the cortical structure of PT and IFGp plays in linguistic tone processing.

In conclusion, the present results indicate that individual differences in the cortical structure of left PT as well as left IFGp might be related to word tone processing in native speakers. The role of these brain regions seems to be modulated by the presence versus absence of lexical content in incoming speech, suggesting that the cortical thickness of left PT might influence processing when tones are accessed as part of stored representations for familiar speech sound sequences, potentially corresponding to whole-word forms, whereas the cortical thickness of left IFGp might play a greater role during rule-based processing of regularly inflected items.

4. Method

4.1. Participants

As the research question of the present study focused on the relationship between the degree of predictive tone processing and cortical structure, participants who did not show a response time advantage for validly cued suffixes, indicating that they did not use tones predictively during the experimental task, were excluded from the present analysis ($n = 3$ in the real word experiment and $n = 2$ in the pseudoword experiment). Thus, data obtained with 15 native speakers of Central Swedish (7 females, mean age: 25.4 years, $SD = 5.2$) participating in the real word experiment, as well as data from 17 native speakers of Central Swedish (11 females, mean age: 24.9 years, $SD = 4.0$) participating in the pseudoword experiment were analysed.

4.2. Stimuli and procedure

There were two experimental conditions, valid and invalid tone-suffix association. In the valid condition, word accents on the stem validly cued their associated suffix, i.e. accent 1 was followed by the singular suffix and accent 2 was followed by the plural suffix. In the invalid condition, the stem tone-suffix combinations were invalid, i.e. accent 2 on the stem preceded the singular suffix and accent 1 preceded the plural suffix. The word stem was always a monosyllabic Swedish word in the real word experiment (e.g. Valid: *hatt*_{Accent 1} + *en*, 'hat + sg'; Invalid: **hatt*_{Accent 1} + *ar*, 'hat + pl'), and a meaningless, but phonotactically legal, pseudoword (pseudoword stem + suffix) in the pseudoword experiment (e.g. Valid: *kvut*_{Accent 1} + *en*, 'kvut + sg'; Invalid: **kvut*_{Accent 1} + *ar*, 'kvut + pl'). Target words were placed in carrier sentences with identical structure, e.g. *Kurt fick [target word] till jul*, 'Kurt got [target word] for Christmas'. Stimulus material preparation involved the same recording and splicing procedures for both experiments. Sentences were presented auditorily and participants' task was to decide, as quickly as possible, if the person mentioned in the sentence received one or many things, by pressing a button. Response times were measured from suffix onset. In the real word experiment, there were 60 different target words (30 different lexical words presented once in singular, once in plural) in each of the two experimental conditions. In the pseudoword experiment, each of the experimental conditions consisted of 80 different target words (40 lexical words presented once in singular, once in plural). For more details, see Roll et al. (2015) and Söderström et al. (2017).

4.3. MRI

T1-weighted MPRAGE MRI scans were collected on Magnetom

Siemens 3T Skyra (real word experiment) and Prisma (pseudoword experiment) systems (TR/TE/TI 1900/2.54/900 ms, 1 mm³ isovoxel, 256 × 256 matrix, 176 slices, flip angle 9°). Cortical reconstruction and volumetric segmentation were performed with the Freesurfer image analysis suite (Dale, Fischl, & Sereno, 1999). Masks of PT and IFGp were taken from the probabilistic Harvard-Oxford Cortical Structure Atlas in FMRIB Software Library thresholded at 20 and 15, respectively. The value in a point in the atlas corresponds to the probability of that point being included in the anatomical structure of interest based on expert raters in studies involving cortical parcellation (Desikan et al., 2006). Mask thresholds were chosen generously to account for individual differences in patterns of cerebral gyri (see Figs. 1 and 2). Mean cortical thickness of the PT and IFGp masks were extracted using Freesurfer tool *mri_segstats*.

4.4. Response times

Correlation analyses were performed relating individual participants' mean cortical thickness values in the left PT as well as in the left IFGp masks to their response time (RT) differences between validly versus invalidly cued suffixes. Results were statistically evaluated using one-tailed t-tests with Bonferroni-corrected p-values for multiple comparisons. Invalid minus valid RT difference was also correlated with the right PT and right IFGp in a separate analysis.

Statement of significance

We found a relationship between the cortical thickness of specific language-related brain areas and the performance of certain associated native language processing skills. Whole word access versus decompositional processing seems to modulate involvement of the left planum temporale and IFG pars opercularis, in line with dual-route models of morphosyntactic processing.

Acknowledgements

This work was supported by Knut and Alice Wallenberg Foundation (grant number 2014.0139); Marcus and Amalia Wallenberg Foundation (grant number 2014.0039); and the Swedish Research Council (grant number 2011-2284). We would like to thank Jimmy Lätt for his help with data analysis.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bandl.2017.12.001>.

References

- Bozic, M., Fonteneau, E., Su, L., & Marslen-Wilson, W. D. (2015). Grammatical analysis as a distributed neurobiological function. *Human Brain Mapping*, 36, 1190–1201. <http://dx.doi.org/10.1002/hbm.22696>.
- Chang, E. F., Rieger, J., Johnson, K. D., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in the human superior temporal gyrus. *Nature Neuroscience*, 13(11), 1428–1432. <http://dx.doi.org/10.1038/nn.2641>.
- Chiarello, C., Vazquez, D., Felton, A., & McDowell, A. (2016). Structural asymmetry of the human cerebral cortex: Regional and between-subject variability of surface area, cortical thickness, and local gyrification. *Neuropsychologia*, 93, 365–379. <http://dx.doi.org/10.1016/j.neuropsychologia.2016.01.012>.
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis I: Segmentation and surface reconstruction. *NeuroImage*, 9(2), 179–194. <http://dx.doi.org/10.1006/nimg.1998.0395>.
- Desikan, R. S., Ségonne, 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, 968–980. <http://dx.doi.org/10.1016/j.neuroimage.2006.01.021>.
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, 97(20), 11050–11055.
- Foster, N. E., & Zatorre, R. J. (2010). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage*, 53, 26–36. <http://dx.doi.org/10.1016/j.neuroimage.2010.01.021>.

- 1016/j.neuroimage.2010.06.042.
- Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, 4, 416.
- Harasty, J., Seldon, H., Chan, P., Halliday, G., & Harding, A. (2003). The left human speech-processing cortex is thinner but longer than the right. *Laterality*, 8(3), 247. <http://dx.doi.org/10.1080/13576500244000175>.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *The Journal of Neuroscience*, 23(29), 9541–9546.
- Karama, S., Ad-dab'bagh, Y., Haier, R., Deary, I., Lyttelton, O., Lepage, C., & Evans, A. (2009). Positive association between cognitive ability and cortical thickness in a representative US sample of healthy 6 to 18 year-olds. *Intelligence*, 37(2), 145–155. <http://dx.doi.org/10.1016/j.intell.2008.09.006>.
- Keller, S., Crow, T., Foundas, A., Amunts, K., & Roberts, N. (2009). Broca's area: Nomenclature, anatomy, typology and asymmetry. *Brain and Language*, 109(1), 29–48. <http://dx.doi.org/10.1016/j.bandl.2008.11.005>.
- Lehtonen, M., Niska, H., Wandt, E., Niemi, J., & Laine, M. (2006). Recognition of inflected words in a morphologically limited language: Frequency effects in monolinguals and bilinguals. *Journal of Psycholinguistic Research*, 35(2), 121–146. <http://dx.doi.org/10.1007/s10936-005-9008-1>.
- Liem, F., Hürschler, M., Jäncke, L., & Meyer, M. (2014). On the planum temporale lateralization in suprasegmental speech perception: Evidence from a study investigating behavior, structure, and function. *Human Brain Mapping*, 35(4), 1779–1789. <http://dx.doi.org/10.1002/hbm.22291>.
- Liem, F., Zaehle, T., Burkhard, A., Jäncke, L., & Meyer, M. (2012). Cortical thickness of supratemporal plane predicts auditory N1 amplitude. *NeuroReport*, 23, 1026–1030. <http://dx.doi.org/10.1097/WNR.0b013e32835abc5c>.
- Meyer, M., Liem, F., Hirsiger, S., Jäncke, L., & Hänggi, J. (2014). Cortical surface area and cortical thickness demonstrate differential structural asymmetry in auditory-related areas of the human cortex. *Cerebral Cortex*, 24, 2541–2552. <http://dx.doi.org/10.1093/cercor/bht094>.
- Moen, I. (1993). Functional lateralization of the perception of Norwegian word tones: Evidence from a dichotic listening experiment. *Brain and Language*, 44, 400–413. <http://dx.doi.org/10.1006/brln.1993.1024>.
- 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. <http://dx.doi.org/10.1093/cercor/bhp026>.
- Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6, 456–463. [http://dx.doi.org/10.1016/S1364-6613\(02\)01990-3](http://dx.doi.org/10.1016/S1364-6613(02)01990-3).
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, 41(1), 245–255. [http://dx.doi.org/10.1016/S0167-6393\(02\)00107-3](http://dx.doi.org/10.1016/S0167-6393(02)00107-3).
- Rakic, P. (1988). Specification of cerebral cortical areas. *Science*, 241(4862), 170–176.
- Riad, T. (2014). *The phonology of Swedish*. Oxford: Oxford University Press.
- Roll, M., Söderström, P., Mannfolk, P., Shtyrov, Y., Johansson, M., van Westen, D., & Horne, M. (2015). Word tones cueing morphosyntactic structure: Neuroanatomical substrates and activation time course assessed by EEG-fMRI. *Brain and Language*, 150, 14–21. <http://dx.doi.org/10.1016/j.bandl.2015.07.009>.
- Söderström, P., Horne, M., Frid, J., & Roll, M. (2016). Pre-activation negativity (PrAN) in brain potentials to unfolding words. *Frontiers in Human Neuroscience*, 10. <http://dx.doi.org/10.3389/fnhum.2016.00512>.
- Söderström, P., Horne, M., Mannfolk, P., van Westen, D., & Roll, M. (2017). Tone-grammar association within words: Concurrent ERP and fMRI show rapid neural pre-activation and involvement of left inferior frontal gyrus in pseudoword processing. *Brain and Language*, 174, 119–126. <http://dx.doi.org/10.1016/j.bandl.2017.08.004>.
- Tyler, L. K., Stamatakis, E. A., Post, B., Randall, B., & Marslen-Wilson, W. D. (2005). Temporal and frontal systems in speech comprehension: An fMRI study of past tense processing. *Neuropsychologia*, 43, 1963–1974. <http://dx.doi.org/10.1016/j.neuropsychologia.2005.03.008>.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92, 231–270. <http://dx.doi.org/10.1016/j.cognition.2003.10.008>.
- 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, 828–836. <http://dx.doi.org/10.1093/cercor/bhm115>.
- Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., ... Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping*, 27, 173–183. <http://dx.doi.org/10.1002/hbm.20176>.

Supplementary material

1. Electrophysiological measures and cortical thickness correlations

Individual electrophysiological data collected in Roll et al. (2015) and Söderström et al. (2017) were related to cortical thickness measures of bilateral planum temporale (PT) and pars opercularis of the inferior frontal gyrus (IFGpo) in the same participants.

1.1 Real word experiment

Roll et al. (2015) reported an increased left-lateralized pre-activation negativity for the word accent with greater predictive value (Accent 1), between 136 and 280 ms following stem tone onset. To quantify neural activity, global root mean squares (gRMS) were calculated for individual ERPs (Roll et al., 2015). The average 136-peak found for Accent 1 was correlated with cortical thickness in bilateral PT and IFGpo in the present analysis. Results indicated a positive association between cortical thickness in left PT and the pre-activation effect ($r = .553, p = .040$, uncorrected for multiple comparisons). No further significant correlations were found (right PT: $r = .287, p = .320$; left IFGpo: $r = .467, p = .092$; right IFGpo: $r = -.145, p = .621$).

1.2 Pseudoword experiment

Söderström et al. (2017) observed an increased pre-activation negativity for the more predictive Accent 1 on pseudowords. The negativity was characterized by a more left posterior distribution 70-170 ms following tone onset and with a more left frontal distribution between 270 and 390 ms. No gRMS data was calculated in that study; therefore, correlation analysis was performed by relating individual ERPs corresponding to the pre-activation effect to cortical thickness in bilateral PT and IFGpo. No significant correlations were obtained (Negativity at 70-170 ms with left PT: $r = .115, p = .707$; right PT: $r = .405, p = .169$; left IFGpo: $r = .377, p = .204$; right IFGpo: $r = -.047, p = .880$. Negativity at 270-390 ms with left PT: $r = .139, p = .650$; right PT: $r = .135, p = .657$; left IFGpo: $r = .028, p = .928$; right IFGpo: $r = -.071, p = .818$).

2. Cortical surface area analysis

Mean cortical surface area measures were extracted from cortical ROIs from the Freesurfer Destrieux (aparc.a2009s) atlas (Fischl et al., 2004) and were analysed in the same way as the cortical thickness data, described in the Method section.

2.1 Results

2.1.1 Real word experiment

There was no significant correlation between response times (RTs) and cortical surface area (left PT: $r = .059, p = .833$; right PT: $r = -.207, p = .459$; left IFGpo: $r = .206, p = .462$; right IFGpo: $r = -.024, p = .934$).

2.1.2 Pseudoword experiment

Cortical surface area showed a significant negative correlation with RTs in left PT ($r = -.517$, $p = .034$). No other significant correlation was obtained (right PT: $r = -.433$, $p = .082$; left IFGpo: $r = .046$, $p = .861$; right IFGpo: $r = -.081$, $p = .757$).

3. Supplementary response time analyses

3.1 Real word experiment

Based on their mean cortical thickness in the left PT, participants were divided into a relatively thinner (2.117 – 2.465 mm, $n = 7$) and a relatively thicker cortex group (2.590 – 2.912 mm, $n = 7$). An independent samples t-test showed that RTs for invalidly cued suffixes were not significantly faster ($t(12) = 1.255$, $p = .117$) in the thicker cortex group ($M = 600$ ms, $SD = 102$) than in the thinner cortex group ($M = 718$ ms, $SD = 227$). The relative response time increase from validly to invalidly cued suffixes was significantly greater ($t(12) = 2.393$, $p = .017$) in the thicker cortex group ($M = 96$ ms, $SD = 35$) as compared to the thinner cortex group ($M = 54$ ms, $SD = 30$).

An independent samples t-test (two-tailed) was also conducted to compare average cortical thickness of left IFGpo across the previously created two groups. Results showed that the thicker cortex group on average did not have significantly thicker cortex ($t(12) = 1.132$, $p = .284$) in the left IFGpo ($M = 2.697$, $SD = 0.101$) than the thinner cortex group ($M = 2.614$, $SD = 0.164$).

3.2 Pseudoword experiment

Based on participants' mean cortical thickness in the left IFGpo, a relatively thinner cortex group (2.426 – 2.742 mm, $n = 8$) and a relatively thicker cortex group (2.758 – 2.992 mm, $n = 8$) were created. An independent samples t-test indicated that RTs for invalidly cued suffixes were marginally faster ($t(14) = 1.672$, $p = .059$) in the thicker cortex group ($M = 880$ ms, $SD = 166$) than in the thinner cortex group ($M = 1068$ ms, $SD = 272$). The relative response time increase from validly to invalidly cued suffixes was marginally larger ($t(14) = 1.693$, $p = .058$) in the thicker cortex group ($M = 117$ ms, $SD = 63$) as compared to the thinner cortex group ($M = 71$ ms, $SD = 44$).

Results of an independent samples t-test (two-tailed) comparing average cortical thickness of left PT across the previously created two groups indicated that the thicker cortex group had marginally ($t(14) = 2.111$, $p = .054$) thicker cortex in the left PT ($M = 2.794$, $SD = 0.195$) relative to the thinner cortex group ($M = 2.602$, $SD = 0.166$).

4. Age effects on cortical thickness

To check for potential effects of individual differences in participants' age on the obtained results, cortical thickness of left PT as well as left IFGpo was correlated with age. No significant correlations were found between age and cortical thickness of left PT ($r = -.339$, $p = .184$) or IFGpo ($r = -.219$, $p = .398$) in the pseudoword experiment. In the real word experiment, age did not correlate with cortical thickness of left IFGpo ($r = -.078$, $p = .783$), but a marginally significant negative correlation was found with left PT ($r = -.463$, $p = .082$).

To further examine whether age-related thinning of PT cortex rather than tone-induced suffix processing would better explain the obtained results in the real word experiment, multiple regression analysis was conducted using the predictors age and RT difference, with cortical thickness of left PT as the dependent variable. Results showed that the two predictors explained 42% of the variance ($R^2 = .423$, $F(2, 12) = 4.40$, $p = .037$), with RT difference being a significant predictor ($\beta = .472$, $p = .030$) and age a marginally significant one ($\beta = -.345$, $p = .077$). Therefore, despite the fact that data from the real word experiment seem to reflect some degree of cortical thinning associated with increasing age, the observed association between tone-processing and cortical thickness of left PT nevertheless still holds.

References

Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D., & ... Dale, M. A. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, 14(1), 11-22. doi:10.1093/cercor/bhg087

Table 1

Mean values and standard deviations (SD) for cortical thickness

	Real word experiment		Pseudoword experiment	
	Mean cortical thickness (mm)	SD	Mean cortical thickness (mm)	SD
Left PT	2,5133	0,2134	2,6863	0,2004
Right PT	2,5845	0,1555	2,7654	0,2216
Left IFGpo	2,6570	0,1328	2,7402	0,1448
Right IFGpo	2,6375	0,1179	2,6869	0,1860

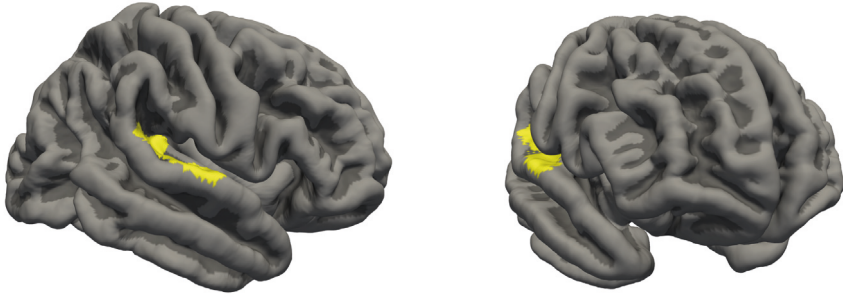


Fig. 1. Right planum temporale presented on the pial surface of Freesurfer common space object fsaverage.

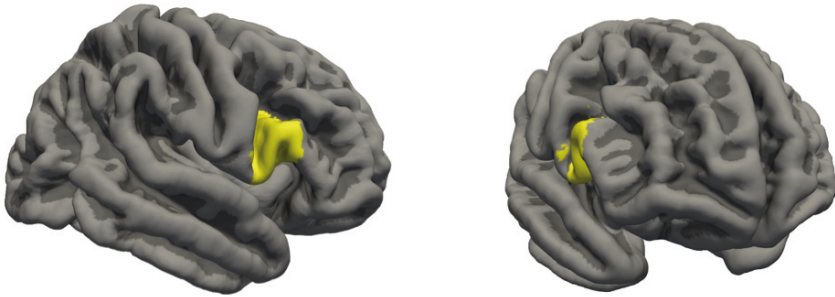
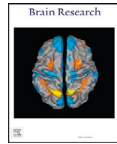


Fig. 2. Right IFG pars opercularis presented on the pial surface of Freesurfer common space object fsaverage.

Study II





Research report

Cortical thickness and surface area of left anterior temporal areas affects processing of phonological cues to morphosyntax



Mikael Novén*, Andrea Schremm, Merle Horne, Mikael Roll

Centre for Languages and Literature, Lund University, Box 201 221 00, Lund, Sweden

HIGHLIGHTS

- Swedish word accents can be used as a measure of perceptual phonological proficiency.
- Cortical thickness and surface area correlate with perceptual phonological proficiency.
- Swedish word accents influence word and phrase recognition in the ventral speech processing stream.

ARTICLE INFO

Keywords:
Speech processing
Prosody
Word accent
Morphology
Cortical thickness
Cortical surface area

ABSTRACT

Lack of methods to experimentally assess the perceptual processing of sound features and allow one to measure differences in phonological proficiency has been a limitation for speech processing studies in native speakers. Tonal features associated with Swedish word-stems, word accents, which cue grammatical suffixes, constitute, however, such sound features that can be exploited to generate measures of reliance on morphosyntactically relevant phonological information during word processing. Specifically, there is a natural variance between native speakers in response time (RT) difference between phonologically valid and invalid word accent-suffix combinations that can be used to quantify perceptual phonological proficiency. This study uses ultra-high field magnetic resonance imaging (MRI) to investigate word accents as phonological cues to morphosyntactic meaning. The study adds to the understanding of the neural basis for both morphosyntactically relevant phonological cues by reporting correlations between differences in listeners' RT for validly and invalidly cued suffixes and cortical thickness in left anterior and middle temporal gyrus, and the left anterior superior temporal sulcus as well as cortical surface area in the left middle and inferior temporal gyri. The cortical areas studied are known constituents of the ventral speech processing stream, necessary for word and phrase recognition.

1. Introduction

In order to understand spoken utterances, the brain has to decode the speech signal according to the phonological rules of a language. Such perceptual phonological processing has been located to the ventral stream of auditory processing running anteriorly along the temporal lobe from the primary auditory cortex in native speakers (DeWitt and Rauschecker, 2012; Hickok and Poeppel, 2004). Moreover, the ventral stream has been shown to be more activated when perceiving normal sentences as compared to sentences consisting of pseudowords, while the dorsal stream has been reported to be active during word repetition (Saur et al., 2008). The anterior temporal lobe as part of the ventral stream has shown activation during phrase building and local syntactic violations, such as incomplete prepositional phrases (Friederici, 2012; Friederici and Gierhan, 2013; Friederici et al., 2003). In order to

investigate the role of suprasegmental features associated with morphosyntactic meaning, finding measures of native speaker proficiency in utilizing these prosodic features is warranted. However, native speaker proficiency in phonological perception and processing has been difficult to measure as it is highly automated and most phonological suprasegmental features are either obligatory present for an utterance to be meaningful (e.g. word tones in Mandarin) or mainly inform pragmatic or syntactic inference (e.g. focus, emotional prosody, and boundary tones), see for instance (Grandjean, 2020; Roll and Horne, 2011). Therefore, studies on the neuroanatomical basis for phonological proficiency have mainly dealt with second language learning (Golestani, 2012; Golestani et al., 2007, 2002; Sebastián-Gallés et al., 2012; Wong et al., 2011) or production in L1 (Schwartz et al., 2012). Both of these modalities of phonological proficiency are more prone to draw on the dorsal stream, since it is involved integration of auditory

* Corresponding author.

E-mail address: mikael.noven@ling.lu.se (M. Novén).

and motor information when perceiving new or difficult speech sounds (DeWitt and Rauschecker, 2012; Golestani, 2015; López-Barroso and de Diego-Balaguer, 2017). If there was a way to measure phonological perceptible proficiency in normal functioning native speakers, it would be possible to investigate the relation between perceptual phonological proficiency and ventral stream areas involved in morphosyntax processing. Swedish offers such an opportunity in the form of two stem-tone cues to upcoming inflectional suffixes called ‘word accents’ (Bruce, 1977; Riad, 2014; Roll et al., 2013, 2015; Söderström et al., 2017a, 2017b). Word accents are realized on the word stem, but are conditioned by suffixes. Therefore, listeners use them as predictive cues to which suffix a word is going to end with. In this way, word accents facilitate, but are not essential for, word interpretation. Thus, we expect to find a natural variance of dependency on these tonal cues within the population of native speakers. This variance can be quantified in relation to the extent to which listeners’ speed in deciding which suffix a word has is affected by the validity (correctness) of the stem-tone-suffix combinations. In other words, the impact of the word accents as phonological cues can be thought of as a measure of perceptual phonological proficiency. Previous studies have found electrophysiological (Roll, 2015; Roll et al., 2010, 2013; Söderström et al., 2017b), haemodynamic (Roll et al., 2015; Söderström et al., 2017a) and cortico-anatomical (Schremm et al., 2018) correlates of this tone-suffix association. To investigate the neural basis for Swedish word accents as perceptual cues to morphosyntax in more detail, ultra-high field (7T) MRI was used in the present study to generate a higher tissue contrast as compared to previous studies in order to investigate cortical areas where phonological information can be assumed to impact the processing of grammatically inflected words.

1.1. Swedish word accents

Swedish has two word accents called accent 1 and accent 2. Which accent is associated with a word-stem, exemplified in Fig. 1 with ‘fläck’ (spot), depends on how the word ends. When inflected with the singular definite suffix ‘-en’ in ‘fläcken’ (the spot), the word-stem carries accent 1, while in ‘fläckar’ (spots), inflected with indefinite plural suffix -ar, it carries accent 2. Accent 1 is a stronger predictor of suffixes than accent 2 since it is associated with on average 11 times fewer word endings than accent 2 (Söderström et al., 2016). The word accents do not express lexical or morphosyntactic information on their own and thus a listener who only hears the word stem ‘fläck’ (spot) cannot determine its inflected form. This implies that while word accents cue word endings,

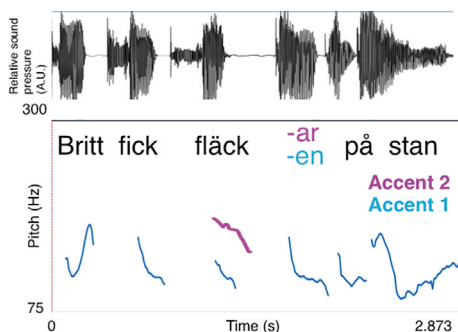


Fig. 1. Waveform (top) and fundamental frequency (pitch) contour (bottom) of the example sentence *Britt fick fläck-ar-en på stan* ‘Britt got stains/the stain in town’ with ‘stain’ being the target word stem. The valid accent 2-plural suffix -ar combination and accent 1-definite singular suffix -en are indicated by color coding.

the impact that they have on the actual suffix processing can only be determined when the suffix itself is processed. As a consequence, in experimental mismatch paradigms where word-stem tones are combined correctly or incorrectly with a given suffix on a test word, direct comparisons between invalidly and validly cued suffixes for each stem tone separately are conflated with suffix processing. Therefore, it is necessary to compare averaged response times to accent-suffix combinations for both word accents in order to obtain a measure of the perceptual phonological proficiency, i.e. the use of the word accents as phonological cues to morphosyntactic processing.

In Central Swedish, the dialect used in this study, accent 1 is realized as a relatively low tone, while accent 2 is realized as a high tone (see Fig. 1) but, importantly, the predictive value of word accents in Central Swedish is preserved in a dialect where the acoustic realizations of the accents are reversed (Roll, 2015). Since accent 1 is a stronger predictor, there is a predictive asymmetry between word accents, reflected in shorter response times for validly cued accent 1 suffixes than for accent 2 suffixes (Söderström et al., 2012) as well as in the increase in an early event-related potential (ERP) known as pre-activation negativity (PrAN) for accent 1 relative to accent 2. PrAN correlates with the predictive value of the phonemes of the word beginning, including word accents (Söderström et al., 2016; Roll et al., 2017). Moreover, ERP studies have shown that suffixes preceded by an invalid word accent cue in Swedish elicit a P600 (Roll et al., 2010, 2013, 2015; Roll, 2015), an ERP component indexing morphological and syntactic errors and repair, ambiguities, and complexity. The role of word accents in morphosyntactic processing is supported by findings of a left-anterior negativity (LAN) in pseudoword experiments in Swedish native speakers (Söderström et al., 2017a) and in learners of Swedish that have been trained in the accent-suffix relation (Hed et al., 2019) as the LAN component is elicited by increased morphological processing (Penke et al., 1997; Weyerts et al., 1997) as well as morphological rule application (Krott and Lebib, 2013). Taken together, previous studies have shown that word accents are morphologically conditioned and that this relation is used by listeners to facilitate morphological and morphosyntactic processing during speech comprehension.

Experiments using functional magnetic resonance imaging (fMRI) have shown effects of word accent processing in left superior temporal areas as well as the temporal pole (parts of the ventral stream) and left IFG (Roll et al., 2015; Söderström, et al., 2017b). Response time advantage for validly over invalidly cued suffixes has been seen to correlate positively with cortical thickness in left planum temporale for real words but in left inferior frontal gyrus pars opercularis for pseudowords (Schremm et al., 2018). Schremm et al. (2018) used predefined ROIs based on results from previous studies and focused on the role of tone in assumed whole-word form processing for real words. The present study expands the inquiry to a spatially unbiased whole-brain correlation analysis between cortical thickness or surface area and dependency on word accents as phonological cues to morphosyntax. This method could possibly lead to discovery of correlations in areas within the ventral processing stream of speech suggested to be causally involved in word comprehension (DeWitt and Rauschecker, 2012, 2016) and in local phrase structure processing (Friederici et al., 2006, 2003). Ventral stream areas would be expected to be implied for measures of phonological proficiency based on processing of word accents in native speakers of Swedish since word accents have been found to influence processing at the word-level. Dorsal stream areas would not be expected to play any considerable role in contributing to this kind of proficiency since the dorsal stream is assumed to be mostly involved in processing novel speech sounds (DeWitt and Rauschecker, 2012; Golestani, 2015; López-Barroso and de Diego-Balaguer, 2017). In the present study, all of the nouns used as targets in the stimulus material are frequent Swedish words, and therefore not likely to be impacted by dorsal stream areas in their processing.

1.2. The present study

The aim of the present study is thus to investigate the correlations between cerebral cortical thickness, surface area and brain activity on the impact of word accent cues to the processing of grammatical number suffixes. This is the first study on the neural correlates of word accent processing using ultra-high field – 7T – MRI (providing increased signal-to-noise ratio and tissue contrast (Duyn, 2012)). Swedish word accents' role in cueing morphosyntactic information leads us to hypothesise a correlation of both cortical thickness and surface area with phonological proficiency in anterior temporal – ventral stream – areas. This hypothesis is in turn based on the radial unit hypothesis which states that the cortex is built up of functionally specialized cortical columns, created during embryonic development (Rakic, 1988). The cortical surface area is, according to the hypothesis, determined by the number of the columns, while the thickness of the cortex is determined by the number and size of cells within the columns. This is corroborated by the independence between cortical surface area and thickness (Panizzon et al., 2009; Meyer et al., 2013; Vuoksima et al., 2015) also explaining why we investigate both cortical thickness and surface area as possible correlates of phonological proficiency.

The structural measurements in the present study were carried out in connection with an fMRI experiment where participants were asked to indicate as quickly as possible if the object in a sentence was in the singular (“one”) or plural (“several”) form. The object either carried an invalid or valid word accent-suffix combination as described in Section 1.1. The mean difference between response times from invalid and valid accent-suffix combinations was used as a measure of phonological proficiency in whole-brain correlation analyses. The behavioral data from the fMRI experiment will also be analyzed and discussed here, but not the fMRI data, since the strategy for ensuring that the participants heard the stimuli introduced severe imaging artefacts. A description of the fMRI sequence can be found in [supplementary material](#). Participants were expected to respond faster and produce fewer errors when the accent-suffix combination was phonologically valid than when it was invalid. Correlations between cortical thickness or surface area and phonological proficiency could indicate that long term reliance on phonological cues to inflections have shaped the cortex as regards cellular composition and organization to allow comprehenders to readily use these cues in their speech comprehension.

2. Results

2.1. Behavior

A summary of the behavioral results from the fMRI experiment is found in [Table 1](#). Responses were faster and more correct when the accent-suffix combinations were valid rather than invalid. Because the suffixes for a given word are associated with different accents such that the valid suffix for accent 1 is the same as the invalid suffix for accent 2 and vice versa, one cannot draw conclusions as regards their impact on whole word processing based on the word-accent suffix combination for each word accent separately. An upper-tailed paired-sample *t*-test revealed significantly longer response times for invalidly than for validly cued suffixes $t(23) = 3.21, p = 0.0020$, likely indicating correction of

wrong suffix predictions based on the stem tone. Accuracy was also significantly impacted by invalid accent-suffix associations $t(23) = 3.08, p = 0.00263, M_{\text{difference}} = 0.0215$. However, accuracy for the invalidly cued suffixes was still very high. Neither the response time difference or accuracy difference correlated with age (Pearson $r = 0.121, p = 0.572$; Pearson $r = -0.229, p = 0.281$, respectively). Participant gender did not affect response time or accuracy difference ($t(18.2) = -0.232, p = 0.819, M_{\text{male}} = 106 \text{ ms}, M_{\text{female}} = 118 \text{ ms}$; $t(20.9) = -1.10, p = 0.284, M_{\text{male}} = 0.057, M_{\text{female}} = 0.038$, respectively).

2.2. Cortical thickness and surface area

Cortical thickness in a cluster expanding over left anterior and middle temporal gyrus (MTG), superior temporal sulcus, anterior middle temporal sulcus and small parts of inferior temporal gyrus (ITG) correlated positively with response time difference between invalidly and validly cued suffixes averaged for both accents, as shown in [Fig. 2](#). Cortical surface area correlated positively with response time difference in the fMRI experiment in two inferotemporal clusters, one including anterior middle temporal gyrus and extending down to anterior ITG and the other in mid ITG, both shown in [Fig. 3](#). Statistics for max voxels within the clusters are found in [Table 2](#).

3. Discussion

The present study aimed at investigating the impact and neural underpinnings of processing word accent cues to grammatical suffixes on word processing in the ventral stream. We report a correlation between cortical thickness as well as surface area measurements and response time differences in areas within the anterior temporal lobe – part of the ventral auditory processing stream.

Participants' performance in determining whether the target noun was in singular or plural form was significantly impaired by invalid word accent cues. This was evident from significantly longer response times and lower accuracy for invalid vs. valid word accent-suffix combinations. Participants were unevenly affected by mismatches between word accent and suffix as shown as variability in response time differences between invalid and valid word accent-suffix combinations. We take this spread to be a measure of perceptual phonological proficiency within native Swedish speakers and used it to investigate potential corticoanatomical correlates of this proficiency in the processing of words in the ventral stream.

The ventral stream model for speech processing presented by DeWitt and Rauschecker (2012) stipulates a word-form processing gradient going anteriorly from the primary auditory cortex along the superior temporal lobe. Along the gradient, words are built up through detection and combination of acoustic features with other kinds of information. The association between response time advantage for validly cued suffixes and cortical thickness as well as surface area of left hemisphere anterior temporal areas could possibly reflect a role for word accents in the ventral stream. Viewing our results in the light of Hickok and Poeppel's dual-stream model of speech processing, the areas involved are suggested to house phonological networks (mid-post STS), lexical interface (posterior MTG and inferior temporal sulcus), and a

Table 1

Mean response times and proportion of accurate answers (based on suffix) for each condition and mean difference between invalid and valid combinations. Max accuracy is 1. Standard deviations within parentheses.

Validity	Valid		Invalid		Mean Invalid-Valid
	1 Singular	2 Plural	1 Plural	2 Singular	
Response time	2720 (± 381) ms	2930 (± 376) ms	2800 (± 426) ms	2790 (± 458) ms	111 (± 77) ms
Accuracy	0.97 (± 0.04)	0.92 (± 0.07)	0.95 (± 0.06)	0.95 (± 0.05)	-0.05 (± 0.06)

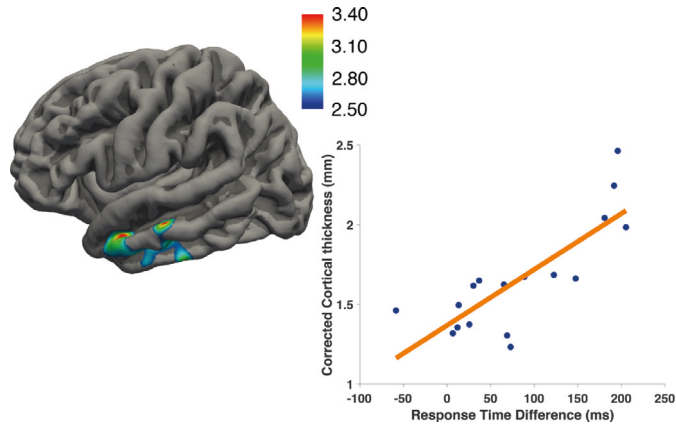


Fig. 2. Cluster of significant correlation between cortical thickness and response time difference between invalidly and validly cued number suffixes (left), color bar indicating Z score, and plot for the max voxel in the middle temporal cortex (MNI coordinates $-51, 1, -28$).

combinatorial network (anterior MTG and inferior temporal sulcus) (Hickok and Poeppel, 2007). Our findings fit the model well, as word accents facilitate both the lexical semantic and morphosyntactic processing of words. Schremm et al. (2018) found that cortical thickness correlated with response time advantage for valid vs invalid Swedish word accent cues in left planum temporale for real words and left inferior frontal gyrus pars opercularis for pseudowords. While Schremm

et al.'s (2018) study analyzed predefined regions of interest of solely planum temporale and pars opercularis of the inferior frontal gyrus the present study used a whole-brain data-driven analysis and found significant correlations in more anterior areas. As the regions of interest in Schremm et al.'s (2018) study were predominantly from the dorsal stream, there was no interpretation of the results in terms of a dual stream model. However, the auditory ventral stream along the superior

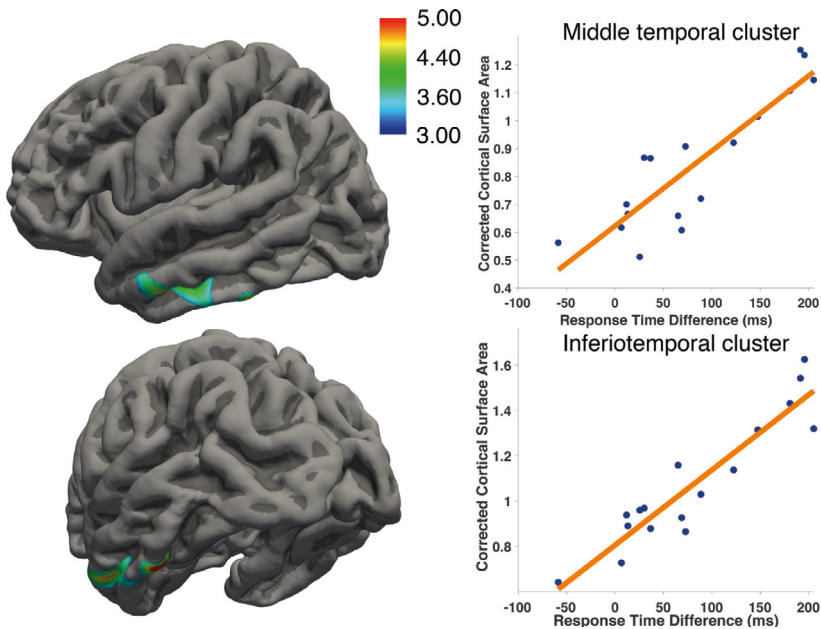


Fig. 3. Clusters of significant correlation between cortical surface area and response time difference between invalidly and validly cued number suffixes from the fMRI experiment, color bar indicating Z score, and plot for the max voxel in the middle and inferior temporal cortex (MNI coordinates $-56, -8, -29$ and $-55, -39, -27$, respectively).

Table 2

Results from cortical thickness and cortical surface area regression analysis on response time difference between valid and invalid word accent-suffix combinations. *cwp* indicates the cluster-wise corrected *p* values.

	Cortical Areas	<i>df</i>	<i>t</i>	<i>r</i>	Peak Coordinate MNI (x, y, z)	<i>cwp</i>
Cortical Thickness	L. anterior superior temporal sulcus, middle temporal gyrus	12	3.15	0.563	(-51, 1, -28)	0.00010
Cortical Surface Area	L. middle temporal gyrus	12	7.46	0.598	(-56, -8, -29)	0.00010
	L. inferior temporal gyrus	12	5.58	0.598	(-55, -39, -27)	0.02400

temporal plane suggested by (DeWitt and Rauschecker, 2013) includes parts of the planum temporale. Cortical thickness in the left inferior frontal gyrus (pars opercularis) correlated with response time advantage in pseudowords. As the meaning of pseudowords, per definition, is unknown, the ventral stream might influence the processing to a less extent, but more research on the difference between effect of word accents on processing real as compared to pseudowords is needed. We suggest that while planum temporale might support memory traces of frequent speech sounds, including word accents, as proposed by Schremm and colleagues anterior STS and MTG could be involved in aiding the assembly of the meaning of the perceived word unit. Because the target words used in the stimuli of this experiment were very frequent Swedish nouns (see Section 5.3), it could be the case that word accents are stored as part of the word form in the proposed auditory word form area (DeWitt and Rauschecker, 2012). The expectation was that greater phonological proficiency would imply more extensive reliance on using the stem tone-suffix relation to interpret the word form. The correlations between reliance on phonological cues and cortical thickness and surface area in ventral stream areas could corroborate such an assumption as it implies that word accents are part of processing word forms. Either long-term effects from learning and integrating tonal cues with morphosyntactic information could cause an increase in cortical thickness over time, or innate variation in the cortical anatomy could perhaps cause an initial sensitivity to tonal information in words.

Friederici (2009) reviewed the white matter pathways in the left hemisphere relevant for language together with functional characterizations of cortical areas connected by the pathways. There are two anatomically and functionally separate ventral streams assumed by this model. One is involved in semantic processes and comprises the triangular and orbital part of the IFG and the mid STG/MTG connected by the extreme capsule. The other is involved in local phrase structure building and consists of the frontal operculum and the anterior STG connected by the uncinata fasciculus. We suggest that the cortical areas reported in our results, all adjacent to the temporal local phrase building ventral stream, possibly facilitate word/phrase processing by computing and transmitting supportive information from the word accents. As the dorsal stream's main role is non-lexical speech perception, such as syllable discrimination, audio-motor integration, and syntactic processing (Friederici, 2017), dorsal stream areas are not expected to be involved in resolving mismatched accent-suffix combinations. This would suggest an additional module to the ventral stream in the model, processing and adding suprasegmental information to semantic and local phrase structure processing. This kind of module would seem to be expected in languages like Swedish where word tones are highly morphologically conditioned. However, more research is needed to further support the existence and characteristics of this kind of prosodic module.

This study has a number of potential limitations that warrant discussion. First, the number of participants with useable brain segmentations was relatively reduced (16 participants). This was due to problems with the magnetic field inhomogeneities present as local signal loss in the images. For most participants these were possible to mitigate using bias field correction softwares but 7T data require close attention to image quality. More method development in pulse sequence design and image processing softwares is needed to further mitigate these issues and make full and reliable use of the greater overall signal strength

and tissue contrasts obtainable by using ultra-high field MRI. Furthermore, the longer inter-stimulus intervals used compared to previous studies necessary for the present fMRI sequence made it possible for the subjects to respond later than otherwise would be the case. However, the response time and accuracy differences followed the same patterns as in previous studies, making it clear that the word accents were in fact useful as cues.

3.1. Conclusions

We conclude that listeners' reliance on word accents as phonological cues to morphosyntactic information is associated with cortical thickness of left anterior middle temporal gyrus, and the left anterior superior temporal sulcus, as well as cortical surface area of left inferior to middle temporal lobe. The two word accents in Central Swedish are associated with – among other grammatical inflections and word endings – singular and plural suffixes. The consequences of violating accent-suffix agreement become clear when comparing response times for validly and invalidly cued suffixes. The degree to which word accents are used in speech perception to process morphosyntactic meaning can be thought of as a measure of phonological proficiency. The correlation between reliance on word accent cues for processing suffix meaning and cortical thickness and surface area of ventral stream areas corroborates the idea that perception of native language phonology and related morphosyntactic features involves greater use of the “meaning”-oriented ventral stream than the sensorimotor-integrative dorsal stream of speech processing.

4. Methods and materials

4.1. Subjects

24 native speakers of central Swedish (12 female and 12 male) aged 20–30 ($M = 24.2$, $SD = 2.90$) participated in the study. None of the participants were compound bilingual or used other languages than Swedish in their family while growing up. The local ethics board approved the study and all participants gave written consent prior to the experiment. All included participants were right-handed as assessed by the Edinburgh handedness index (Oldfield, 1971).

4.2. Audiometry

Fixed-frequency Békésy audiometry was used to estimate subjects' pure tone hearing thresholds for frequencies 250, 500, 1000, 2000, 4000, and 8000 Hz using a GN Otometrics Astera audiometer and a pair of circumaural sound attenuating Sennheiser HDA 200 earphones. The stimuli consisted of pulsed pure tones gated on 250 ms with a 250 ms silent interval between presentations and a 2 dB intensity rate change per second was used. Six reversals were recorded per frequency and the threshold at each test frequency was calculated as the mean of these six reversals. The calibration of this set-up was made using a Brüel and Kjaer 2231 sound level meter with a 4134 microphone in a 4153 artificial ear according to IEC 60318-2 and ISO 389-8 (International Electrotechnical Commission, 1998). All subjects had normal hearing defined as pure-tone hearing thresholds < 20 dB hearing level (ISO, 2004) for all presented frequencies.

4.3. Procedure and stimuli

30 different monosyllabic noun stems were used as target words in carrier sentences of the form 'Britt fick TARGET på stan', 'Britt got TARGET in town', where the recipient and the place were varied – though never for the same target word – but the verb and word order were always the same. The target words were very common: minimum, maximum and mean lemmagram frequency of 0.8, 36.7 and 13.9 token counts per million (Borin et al., 2012). The experimental conditions were valid and invalid tone-suffix combinations, e.g. Valid: $fläck_{\text{accent } 1} + en$ (stain_{accent 1} + singular)/ $fläck_{\text{accent } 2} + ar$ (stain_{accent 2} + plural) and Invalid $fläck_{\text{accent } 2} + en$ (stain_{accent 2} + singular)/ $fläck_{\text{accent } 1} + ar$ (stain_{accent 1} + plural). A male Central Swedish speaker recorded the sentences in an anechoic chamber. Target words were cut between stem and suffix and by combining stem and suffix, 4 variants of each target word were created: accent 1/2 with valid/invalid suffix. This yielded 120 stimulus sentences. Participants were asked to respond by making a button press as quickly as possible as to whether the object in the sentence was “one” or “many” things. Response times were measured from suffix onset. Accuracy measures were calculated as number of correct answers divided by number of sentences for the condition/condition group. The stimuli were presented every 13.3 s, with a jittered stimulus onset in a 500 ms time window, in an introduced silent period in the epi sequence.

4.4. Magnetic resonance imaging

MRI was performed with an actively shielded 7T scanner (Achieva, Philips, Best, Netherlands) equipped with a dual transmit head coil and a 32 channel receive phased-array head coil (Nova Medical, Wilmington, MA). Dielectric pads were used to reduce the effect of B1-field inhomogeneities (Teuwwisse et al., 2012). Two 3D magnetization-prepared rapid gradient echo (MPRAGE) sequences were acquired with a field of view of $200 \times 252 \times 190 \text{ mm}^3$ and a $1 \times 1 \times 1 \text{ mm}$ voxel size. In order to address inherent inhomogeneities due to the high field strength, one T1 and one proton density (PD) weighted volume were acquired with repetition time (TR) = 5 ms, echo time (TE) = 2 ms, flip angle = 6°, inversion pulse delay = 1200 ms, and TR = 6 ms, TE = 2.5 ms, flip angle = 7°, respectively. This approach is based on the concept of MP2RAGE (Marques et al., 2010). The fMRI sequence used caused artefacts that were too severe for making reliable group analyses. See description and discussion in supplementary materials.

4.5. Cortical thickness and surface area analysis

The image analysis was comprised of a preprocessing step followed by segmentation. For each subject, the PD-weighted volume was registered to the T1-weighted volume. Subsequently, the T1-weighted volume was divided by the PD-weighted and improved nonparametric nonuniform intensity normalization bias field correction was applied (version 2.1.0.post685-g86e09) (Tustison et al., 2010). Data from 8 subjects were discarded from further analysis due to image artefacts.

Cortical reconstruction and volumetric segmentation was performed with the Freesurfer image analysis suite v6.0. The technical details of these procedures and how continuous maps of cortical thickness and surface area are obtained have been described elsewhere (Dale et al., 1999; Fischl and Dale, 2000; Fischl et al., 2002, 1999, 2004; Han et al., 2006; Rosas et al., 2002; Ségonne et al., 2004) A brief summary of how cortical anatomical measurements are estimated and mapped follows.

All Freesurfer surfaces consist of vertices in a triangular grid. Calculated cortical surface areas were mapped onto each vertex as a third of each of the areas of the triangles the vertex was part of. The triangles' area was calculated based on the relative areal expansion of a standardized spherical atlas space as it was mapped to a subject's native space. Cortical thickness data was mapped as the shortest distance between the white matter and pial surface for each vertex of the white

matter surface. Maps of the cortical thickness and surface area data from each subject were transformed into Freesurfer common space, fsaverage, before a 10 mm full width at half maximum Gaussian spatial smoothing kernel was applied.

As brain size is positively correlated with cortical thickness and cortical surface area (Im et al., 2008), estimated intracranial volume (eICV), given from Freesurfer output, was added as a covariate together with age and gender. Regression analyses were performed using Freesurfer tool qdec. Clusters in which cortical thickness or surface area correlated with response time difference between invalidly and validly cued suffixes averaged for both accents underwent comparison to cluster size limits derived from Monte Carlo simulation (6000 permutations)-based multiple comparison correction (Hagler et al., 2006) using a cluster-forming threshold of $p < 0.005$ or $p < 0.001$ for thickness and surface area, respectively, in accordance with suggestions in (Greve and Fischl, 2018). Results were projected onto the pial surface of Freesurfer common space, fsaverage, for display purposes.

Funding

This work was supported by Knut and Alice Wallenberg Foundation (Grant No. 2014.0139) and Marcus and Amalia Wallenberg Foundation (Grant No. 2014.0039).

CRediT authorship contribution statement

Mikael Novén: Conceptualization, Methodology, Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **Andrea Schremm:** Conceptualization, Methodology, Investigation. **Merle Horne:** Conceptualization, Writing - review & editing. **Mikael Roll:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Knut and Alice Wallenberg Foundation (Grant No. 2014.0139) and Marcus and Amalia Wallenberg Foundation (Grant No. 2014.0039). Lund University Bioimaging Center (LBIC), Lund University, Sweden, is gratefully acknowledged for providing experimental resources. We would like to thank an anonymous reviewer for helpful comments which have led to a much improved manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.brainres.2020.147150>.

References

- Borin, L., Forsberg, M., Roxendal, J., 2012. Korp – the corpus infrastructure of Språkbanken. In: Paper presented at the LREC, Istanbul.
- Bruce, G., 1977. Swedish Word Accents in Sentence Perspective. Gleerups, Lund.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. Neuroimage 9 (2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>.
- DeWitt, L., Rauschecker, J.P., 2012. Phoneme and word recognition in the auditory ventral stream. Proc. Natl. Acad. Sci. U. S. A. 109 (8), E505–E514. <https://doi.org/10.1073/pnas.1113427109>.
- DeWitt, L., Rauschecker, J.P., 2013. Wernicke's area revisited: parallel streams and word processing. Brain Lang. 127 (2), 181–191. <https://doi.org/10.1016/j.bandl.2013.09.014>.
- DeWitt, L., Rauschecker, J.P., 2016. Convergent evidence for the causal involvement of

- anterior superior temporal gyrus in auditory single-word comprehension. *Cortex* 77, 164–166. <https://doi.org/10.1016/j.cortex.2015.08.016>.
- 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>.
- Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U.S.A.* 97 (20), 11050–11055.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al., 2002. Whole Brain Segmentation. *Neuron* 33 (3), 341–355. [https://doi.org/10.1016/S0896-6273\(02\)00569-X](https://doi.org/10.1016/S0896-6273(02)00569-X).
- Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis. *Neuroimage* 9 (2), 195–207. <https://doi.org/10.1006/nimg.1998.0396>.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., et al., 2004. Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14 (1), 11–22.
- Friederici, A.D., 2009. Pathways to language: fiber tracts in the human brain. *Trends Cogn. Sci.* 13 (4), 175–181. <https://doi.org/10.1016/j.tics.2009.01.001>.
- Friederici, A.D., 2012. The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.* 16 (5), 262–268. <https://doi.org/10.1016/j.tics.2012.04.001>.
- Friederici, A.D., 2017. Evolution of the neural language network. *Psychonomic Bull. Rev.* 24 (1), 41–47. <https://doi.org/10.3758/s13423-016-1090-x> Retrieved from doi: 10.3758/s13423-016-1090-x.
- Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I., Anwander, A., 2006. The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci.* 103 (7), 2458–2463. <https://doi.org/10.1073/pnas.0509389103>.
- Friederici, A.D., Gierhan, S.M.E., 2013. The language network. *Curr. Opin. Neurobiol.* 23 (2), 250–254. <https://doi.org/10.1016/j.conb.2012.10.002>.
- Friederici, A.D., Rüschemeyer, S.-A., Hahne, A., Fiebach, C.J., 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb Cortex* 13 (2), 170–177. <https://doi.org/10.1093/cercor/13.2.170>.
- 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. *Int. J. Bilingualism* 18 (1), 6–34. <https://doi.org/10.1177/1367006912456585>.
- Golestani, N., 2015. Neuroimaging of phonetic perception in bilinguals. *Bilingualism: Lang. Cogn.* 19 (4), 674–682. <https://doi.org/10.1017/S1366728915000644>.
- Golestani, N., Molko, N., Dehaene, S., LeBihan, D., Pallier, C., 2007. Brain structure predicts the learning of foreign speech sounds. *Cereb Cortex* 17 (3), 575–582. <https://doi.org/10.1093/cercor/bhk001>.
- 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\(02\)00852-0](https://doi.org/10.1016/S0896-6273(02)00852-0).
- Grandjean, D., 2020. Brain networks of emotional prosody processing. *Emotion Rev.* <https://doi.org/10.1177/1754073919898522>.
- 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>.
- Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., et al., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage* 32 (1), 180–194. <https://doi.org/10.1016/j.neuroimage.2006.02.051>.
- Hed, A., Schremm, A., Horne, M., Roll, M., 2019. Neural correlates of second language acquisition of tone-grammar associations. *Mental Lexicon* 14 (1), 98–123. <https://doi.org/10.1075/ml.17018.hed>.
- 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>.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. [Perspective]. *Nat. Rev. Neurosci.* 8, 393. <https://doi.org/10.1038/nrn2113>.
- 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. *Cereb Cortex* 18 (9), 2181–2191. <https://doi.org/10.1093/cercor/bhm244>.
- International Electrotechnical Commission, I., 1998. *Electroacoustics: Simulators of Human Head and Ear, Part 1: Ear Simulator for the Calibration of Supra-aural Earphones*. IEC, Geneva, Switzerland.
- ISO, 2004. *Int. SO 389-8: Acoustics: Reference Zero for the Calibration of Audiometric Equipment. Part 8: Reference Equivalent Threshold Sound Pressure Levels for Pure Tones and Circumaural Earphones*. International Organization for Standardization, pp. 389–398.
- Krott, A., Lehib, R., 2013. Electrophysiological evidence for a neural substrate of morphological rule application in correct wordforms. *Brain Res.* 1496, 70–83. <https://doi.org/10.1016/j.brainres.2012.12.012>.
- López-Barroso, D., de Diego-Balaguer, R., 2017. Language learning variability within the dorsal and ventral streams as a cue for compensatory mechanisms in aphasia recovery. [Mini Review]. *Front. Hum. Neurosci.* 11 (476). <https://doi.org/10.3389/fnhum.2017.00476>.
- Marques, J.P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., Gruetter, R., 2010. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *Neuroimage* 49 (2), 1271–1281. <https://doi.org/10.1016/j.neuroimage.2009.10.002>.
- 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. *Cereb. Cortex* 24 (10), 2541–2552. <https://doi.org/10.1093/cercor/bht094>.
- 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).
- Panizzon, M.S., Fennema-Notestine, C., Eyer, L.T., Jernigan, T.L., Prom-Wormley, E., Neale, M., et al., 2009. Distinct genetic influences on cortical surface area and cortical thickness. *Cereb. Cortex* 19 (11), 2728–2735. <https://doi.org/10.1093/cercor/bhp026>.
- Penke, M., Weyerts, H., Gross, M., Zander, E., Munte, T.F., Clahsen, H., 1997. How the brain processes complex words: an event-related potential study of German verb inflections. *Brain Res. Cogn. Brain Res.* 6 (1), 37–52.
- Rakic, P., 1988. Specification of cerebral cortical areas. *Science* 241 (4862), 170–176. <https://doi.org/10.1126/science.3291116>.
- Riad, T., 2014. *The Phonology of Swedish*. Oxford University Press.
- Roll, M., 2015. A neurolinguistic study of South Swedish word accents: electrical brain potentials in nouns and verbs. *Nordic J. Ling.* 38 (2), 149–162. <https://doi.org/10.1017/S00332586515000189>.
- Roll, M., Horne, M., 2011. Interaction of right- and left-edge prosodic boundaries in syntactic parsing. *Brain Res.* 1402, 93–100. <https://doi.org/10.1016/j.brainres.2011.06.002>.
- Roll, M., Horne, M., Lindgren, M., 2010. Word accents and morphology—ERPs of Swedish word processing. *Brain Res.* 1330, 114–123. <https://doi.org/10.1016/j.brainres.2010.03.020>.
- Roll, M., Söderström, P., Frid, J., Mannfolk, P., Horne, M., 2017. Forehearing words: Pre-activation of word endings at word onset. *Neurosci. Lett.* 658, 57–61. <https://doi.org/10.1016/j.neulet.2017.08.030>.
- Roll, M., Söderström, P., Horne, M., 2013. Word-stem tones cue suffixes in the brain. *Brain Research* 1520, 116–120. <https://doi.org/10.1016/j.brainres.2013.05.013>.
- Roll, M., Söderström, P., Mannfolk, P., Shtyrov, Y., Johansson, M., van Westen, D., Horne, M., 2015. Word tones cueing morphosyntactic structure: neuroanatomical substrates and activation time-course assessed by EEG and fMRI. *Brain Lang.* 150, 14–21. <https://doi.org/10.1016/j.bandl.2015.07.009>.
- Rosas, H.D., Liu, A.K., Hersch, S., Glessner, M., Ferrante, R.J., Salat, D.H., et al., 2002. Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology* 58 (5), 695–701.
- Saur, D., Kreher, B.W., Schnell, S., Kümmerer, D., Kellmeier, P., Vry, M.-S., et al., 2008. Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci.* 105 (46), 18035–18040. <https://doi.org/10.1073/pnas.0805234105>.
- Schremm, A., Novén, M., Horne, M., Söderström, P., van Westen, D., Roll, M., 2018. Cortical thickness of planum temporale and pars opercularis in native language tone processing. *Brain Lang.* 176, 42–47. <https://doi.org/10.1016/j.bandl.2017.12.001>.
- Schwartz, M.F., Faseyitan, O., Kim, J., Coletti, H.B., 2012. The dorsal stream contribution to phonological retrieval in object naming. *Brain* 135 (12), 3799–3814. <https://doi.org/10.1093/brain/aww300>.
- Sebastián-Gallés, N., Soriano-Mas, C., Baus, C., Díaz, B., Ressel, V., Pallier, C., et al., 2012. Neuroanatomical markers of individual differences in native and non-native vowel perception. *J. Neurolin.* 25 (3), 150–162. <https://doi.org/10.1016/j.jneuroling.2011.11.001>.
- Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22 (3), 1060–1075. <https://doi.org/10.1016/j.neuroimage.2004.03.032>.
- Söderström, P., Horne, M., Frid, J., Roll, M., 2016. Pre-Activation Negativity (PrAN) in brain potentials to unfolding words. [Original Research]. *Front. Hum. Neurosci.* 10 (512). <https://doi.org/10.3389/fnhum.2016.00512>.
- Söderström, P., Horne, M., Mannfolk, P., van Westen, D., Roll, M., 2017a. Tone-grammar association within words: concurrent ERP and fMRI show rapid neural pre-activation and involvement of left inferior frontal gyrus in pseudoword processing. *Brain Lang.* 174, 119–126. <https://doi.org/10.1016/j.bandl.2017.08.004>.
- Söderström, P., Horne, M., Roll, M., 2017b. Stem Tones Pre-activate Suffixes in the Brain. *J. Psycholinguist. Res.* 46 (2), 271–280. <https://doi.org/10.1007/s10936-016-9434-2>.
- Söderström, P., Roll, M., Horne, M., 2012. Processing morphologically conditioned word accents. *Mental Lexicon* 7 (1), 77–89. <https://doi.org/10.1075/ml.7.1.04soe>.
- 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. *Magn. Reson. Med.* 67 (5), 1285–1293. <https://doi.org/10.1002/mrm.23108>.
- 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 Trans. Med. Imaging* 29 (6), 1310–1320. <https://doi.org/10.1109/TMI.2010.2046908>.
- Vuoksimaa, E., Panizzon, M.S., Chen, C.-H., Fiecas, M., Eyer, L.T., Fennema-Notestine, C., et al., 2015. The genetic association between neocortical volume and general cognitive ability is driven by global surface area rather than thickness. *Cereb. Cortex* 25 (8), 2127–2137. <https://doi.org/10.1093/cercor/bhu018>.
- Weyerts, H., Penke, M., Dohrn, U., Clahsen, H., Munte, T.F., 1997. Brain potentials indicate differences between regular and irregular German plurals. *NeuroReport* 8 (4), 957–962.
- 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. *J. Neurosci.* 31 (24), 8780–8785. <https://doi.org/10.1523/jneurosci.0999-11.2011>.

Study III





Short communication

Cortical thickness of Broca's area and right homologue is related to grammar learning aptitude and pitch discrimination proficiency

Mikael Novén^{a,*}, Andrea Schremm^a, Markus Nilsson^b, Merle Horne^a, Mikael Roll^a^a Department of Linguistics and Phonetics, Lund University, 221 00, Sweden^b Lund University, Faculty of Medicine, Department of Clinical Sciences Lund, Radiology, Lund 221 00, Sweden

ARTICLE INFO

Keywords:
 Language learning aptitude
 Cortical thickness
 Broca's area
 Inferior frontal gyrus

ABSTRACT

Aptitude for and proficiency in acquiring new languages varies in the human population but their neural bases are largely unknown. We investigated the influence of cortical thickness on language learning predictors measured by the LLAMA tests and a pitch-change discrimination test. The LLAMA tests are first language-independent assessments of language learning aptitude for vocabulary, phonetic working memory, sound-symbol correspondence (not used in this study), and grammatical inferring. Pitch perception proficiency is known to predict aptitude for learning new phonology. Results show a correlation between scores in a grammatical meaning-inferencing aptitude test and cortical thickness of Broca's area ($r(30) = 0.65$, $p = 0.0202$) and other frontal areas ($r(30) = 0.66$, $p = 0.0137$). Further, a correlation was found between proficiency in discriminating pitch-change direction and cortical thickness of the right Broca homologue ($r(30) = 0.57$, $p = 0.0006$). However, no correlations were found for aptitude for vocabulary learning or phonetic working memory. Results contribute to locating cortical regions important for language-learning aptitude.

1. Introduction

It is becoming increasingly evident that language learning reshapes the brain (Li, Legault, & Litcofsky, 2014). Cortical thickness has been shown to increase in language-related areas as students learn a new language (Mårtensson et al., 2012). However, less is known about how the cortical anatomy prior to training influences the learning of language skills although it has been suggested as a research focus (Golestani, 2012). A better understanding of the extent to which the brain uses similar strategies and neural substrates for native language processing and second language acquisition is also important for the development of language-teaching methods. Measures of cortical anatomy such as cortical thickness are measurable using magnetic resonance imaging (MRI) that produces images with grey/white matter contrast. This study aims to further our understanding of how the brain acquires new language-related knowledge by assessing the relationship between measures of language learning aptitude, pitch discrimination and cortical thickness in a spatially unbiased whole-brain analysis.

Differences in how easily people learn new languages can be assessed in aptitude tests (Carroll & Sapon, 1959; Meara, 2005). The LLAMA tests (Meara, 2005) are native language-independent, computer-based assessments of aptitude for learning foreign languages. The

LLAMA tests consist of four sub-tests, each assessing aptitude for a different language learning modality: written vocabulary (LLAMA B), phonetic memory (LLAMA D), sound-symbol correspondence (LLAMA E), and grammatical inferring (LLAMA F). In studies of L2 attainment, LLAMA test scores have been found to correlate with scores in grammaticality judgment tests (Abrahamsson & Hyltenstam, 2008), morphosyntactic attainment (Granena, 2012), collocation knowledge (Forsberg Lundell & Sandgren, 2013; Granena & Long, 2013), and pronunciation (Granena & Long, 2013). Since the goal of the present study did not involve investigation of orthographic knowledge, we excluded the LLAMA E subtest focusing on sound-symbol correspondence. Suspected cortical areas that influence the LLAMA B scores measuring the associative learning of written words to pictures include bilateral posterior supramarginal gyrus in which the grey matter density has been shown to be proportionally related to size of vocabulary (Lee et al., 2007). In an fMRI experiment, phonological working memory, recruited in the LLAMA D test, has been found to involve the superior temporal gyrus (STG), supplementary motor area, and inferior frontal gyrus (IFG) (Perrachione, Ghosh, Ostrovskaya, Gabrieli, & Kovelman, 2017). Higher LLAMA F scores have been associated with higher white matter connectivity in the left than right tract between Brodmann area (BA) 45, part of Broca's area, and the posterior temporal

* Corresponding author at: Språk- och Litteraturcentrum, Box 201, 221 00 Lund, Sweden.

E-mail addresses: mikael.noven@ling.lu.se (M. Novén), andrea.schremm@ling.lu.se (A. Schremm), markus.nilsson@med.lu.se (M. Nilsson), merle.horne@ling.lu.se (M. Horne), mikael.roll@ling.lu.se (M. Roll).<https://doi.org/10.1016/j.bandl.2018.12.002>

Received 28 January 2018; Received in revised form 23 November 2018; Accepted 12 December 2018

Available online 18 December 2018

0093-934X/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

lobe (Xiang, Dediu, Roberts, Oort, Norris, & Hagoort, 2012). As shown in fMRI and lesion studies, Broca's area in the left IFG (LIFG) modulates grammar processing, with BA 45 more involved in processing local relations, and relating morphology, and morphosyntax to grammatical meaning, and BA 44 more focused on processing syntax (Goucha & Friederici, 2015; Marslen-Wilson & Tyler, 2007). Cortical thickness in LIFG increases when children learn a second language (Klein, Mok, Chen, & Watkins, 2014). Studies assessing the relation between cortical thickness and LLAMA test scores have not been reported prior to this study, despite associations between cortical thickness and other cognitive measures such as general intelligence (Haier, Jung, Yeo, Head, & Alkire, 2004; Karama, Ad-Dab'bagh, Haier, Deary, Lyttelton, Lepage, Brain Development Cooperative Group, 2009; Menary et al., 2013), as well as the known increase in cortical thickness of Broca's area during childhood (Sowell et al., 2004).

Proficient pitch perception is known to facilitate learning to perceive and produce new phonological patterns (Posedel, Emery, Souza, & Fountain, 2011; Slevc & Miyake, 2006), but proficiency in processing pitch is not tested in the LLAMA tests nor in other aptitude tests. Pronunciation proficiency in a learned second language has been seen to be associated with less activity as well as greater gray matter volume in a left hemisphere perisylvian network (Reiterer et al., 2011). Pitch variation can convey aspects of grammar such as morphosyntactic (Roll et al., 2015; Söderström, Horne, & Roll, 2017) and lexical information (Wang, 1973). The ability to discriminate variations in pitch is closely connected to cortical structures in the right hemisphere, involving the homologue of Broca's area as well as auditory cortex, the intraparietal sulcus, anterior middle frontal gyrus, and premotor cortex (Friederici & Alter, 2004; Zatorre, Belin, & Penhune, 2002). Interestingly, persons with amusia have a thicker cortex in frontal and temporal areas of the right hemisphere (Hyde et al., 2007). These results are of interest in investigating aptitude for acquisition of tone languages since language-related tones, like musical tones, are thought to be processed by the same brain areas, at least during early stages of language acquisition (Gosselke Berthelsen, Horne, Brännström, Shtyrov, & Roll, 2018). We tested proficiency in discriminating direction of pitch change, since this is of importance in learning intonation patterns and lexical tones. Prior to testing, subjects' hearing was assessed using fixed-frequency Békésy audiometry in order to be able to exclude participants with hearing deficiencies. Proficiency in pitch-discrimination was assessed using an in-house test developed to measure the minimum difference in pitch needed to judge the direction of a pitch change.

The aim of the present study was to investigate the relationship between cortical thickness and language learning aptitude, assessed using the LLAMA tests B, D and F, as well as proficiency in pitch discrimination, which can be thought to be a sub-component of language learning aptitude. Cortical thickness was measured from T1-weighted MRI images as the shortest distance between the white-gray matter boundary and the gray matter-cerebrospinal fluid boundary using the freely available software suite *Freesurfer*. In order not to impose any bias to any specific cortical area in our analysis, a whole-brain analysis was performed and all significant clusters reported.

2. Results

The score from the LLAMA F grammatical inferencing aptitude test correlated positively with cortical thickness in the left IFG pars triangularis (LIFGpt) part of Broca's area ($r(30) = 0.65$, $p = 0.0202$) and left medial superior frontal gyrus (LMFG) ($r(30) = 0.66$, $p = 0.0137$) (Fig. 1, Table 1A). All reported correlations include age and intracranial volume (ICV) as covariates and were clusterwise corrected for multiple comparisons (p -values computed from the correction method). No significant correlations between cortical thickness and scores from the LLAMA B or D tests were found. Analysis of the pitch discrimination test scores showed that the minimum pitch difference a subject needed to discriminate pitch change direction correlated positively with cortical

thickness in the right inferior frontal gyrus pars triangularis (RIFGpt) and rostral middle frontal gyrus (RMFG) ($r(30) = 0.57$, $p = 0.0006$) (Fig. 2, Table 1B). This indicates that participants with a thicker cortex in the right homologue of Broca's area (RIFGpt) and RMFG performed worse in discriminating pitch change direction.

3. Discussion

We found that cortical thickness in LIFGpt and LMFG correlates with the ease with which a person can identify grammatical form-meaning patterns in a foreign language as measured in the LLAMA F test. LLAMA F involves inferring grammatical meaning associated with linear word-internal morpheme order, classifiers and inflections, all part of the supposed role of LIFGpt (Goucha & Friederici, 2015; Petersson & Hagoort, 2012; Marslen-Wilson & Tyler, 2007). Associating morpho-syntactic form and meaning requires grammatical inferencing which can be assumed to be a general analytical language ability. LMFG is known to be engaged in deductive reasoning (Goel, Gold, Kapur, & Houle, 1997) crucial to the analytical skills required in the LLAMA F test. Results show no connection between cortical thickness and aptitude for the remaining LLAMA subtests (B and D). Cortical thickness can be expected to constitute a reliable predictor of skills that primarily engage a brain area specialized in associating different kinds of information. This finding would suggest that aptitude for grammatical inferencing involves a less distributed neural substratum than aptitude for vocabulary learning or phonetic memory but a more specific substratum where grammatical form-meaning associations are made. One could well expect that understanding auditory linguistic stimuli or learning novel vocabulary rely on different cognitive functions such as attention, relating input to experience and hearing ability. LLAMA B and D scores could perhaps then be better related to the brain using techniques that are capable of detecting patterns in neuronal activity, e.g. resting state fMRI and electroencephalography (EEG). It has been shown that the hemispheric asymmetry in gray-matter volume in the IFGpt correlates with performance on grammaticality judgments (Nauchi & Sakai, 2009). Our results for cortical thickness are also in line with the finding of Xiang et al. (2012) who found significant correlations for LLAMA F score for measures of structural connectivity between BA45, an area that overlaps with IFGpt, and the posterior temporal lobe as well as between BA6 and the posterior temporal lobe. Connectivity of white matter tracts originating from Broca's area has further been found to predict aptitude for artificial grammar learning (Flöel, de Vries, Scholz, Breitenstein, & Johansen-Berg, 2009).

In contrast to the results of the grammatical inferencing task, results from the test of pitch direction discrimination proficiency showed that a greater cortical thickness in RIFG correlated with a decrease in performance in the discrimination task. Unlike grammatical inference, which involves processing many different kinds of parameters associated with linear word-internal morpheme order, classifiers and inflections, discrimination of tonal direction involves decision-making associated with only the parameter of pitch. Cortical thickness in RIFG and the right auditory cortex is negatively correlated with musicality score (Hyde et al., 2007). The indication that a thicker cortex in RIFG is associated with congenital amusia is interpreted by the authors as a result of a less functional neuronal environment due to insufficient pruning of neurites or disturbances to normal neural migration during the development of the cortex. Neuritic pruning and myelination have been suggested as explanations for why thinning of the cortex is associated with a richer acquired vocabulary in children and adolescents (Porter, Collins, Muetzel, Lim, & Luciana, 2011; Sowell et al., 2004). In contrast to the previous findings for amusia, our results for pitch discrimination proficiency do not correlate with cortical thickness in right auditory cortex, but in RIFG. This can be interpreted as indicating that proficiency in judging direction of change in pitch level is specifically associated with the RIFG. The pitch discrimination test included in the present study is intended to provide information that reflects a specific

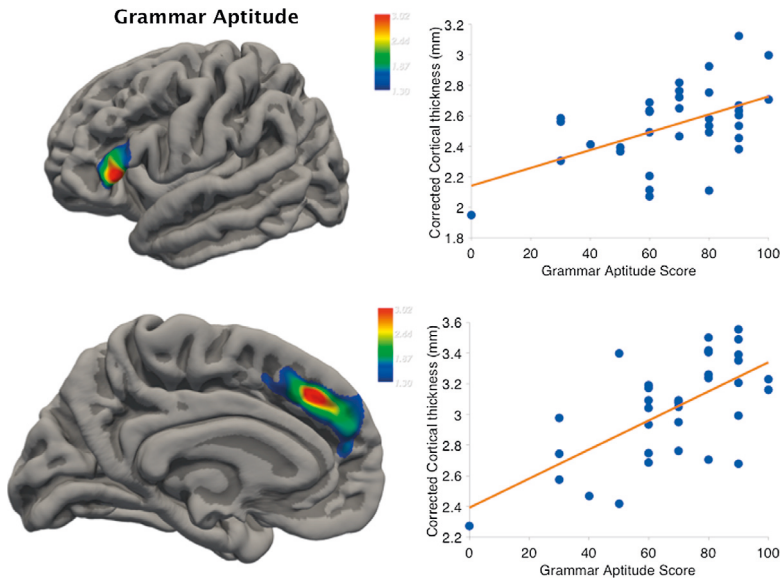


Fig. 1. Left: Significant clusters showing positive correlation between score for grammatical inferring aptitude test and cortical thickness given age and eICV. Color bar indicating Z statistic. Right: Adjusted response plots from max voxel from cluster in left inferior frontal gyrus (top) and left medial frontal gyrus (bottom).

Table 1

A: Statistics for cluster showing significant positive correlation with score for grammatical inferring aptitude test given age and estimated intracranial volume (eICV) as covariates. B: Statistics for cluster showing significant positive partial correlation with pitch discrimination threshold given age and eICV as covariates. Region as defined by aparc atlas. MNI space coordinates. Max r is Pearson's r max value within the significant cluster. CWP is the cluster-wise corrected p-value after Monte Carlo null-Z simulations.

Region	Cluster size (mm ²)	Max Z	X	Y	Z	Max r	CWP
A							
Medial Superiorfrontal	1512.92	4.450	-8	35	36	0.6631	0.01370
IFG pars triangularis	1430.04	4.314	-51	33	-5	0.6544	0.02020
B							
Rostral Middle Frontal Gyrus, IFG pars triangularis	2494.48	3.143	37	39	3	0.5669	0.00060

ability that can be important for learning languages. This would also be a step towards disentangling what in musical ability is associated with aptitude for learning new receptive and productive phonology (Slevc & Miyake, 2006). Future studies focusing on acquisition of tone and intonation will provide important information as to the possible place of a specific pitch processing skill component in the general theory of aptitude for language learning.

Interpreting our results in terms of cognitive load theory (Sweller, 1988) leads us to speculate that performance in tasks involving low cognitive load shows an inverse relation with cortical thickness due to demands on neuronal network efficiency. A thinner cortex could mean a more streamlined design for one type of task or input that can be represented with fewer connections and/or cells. Skills involving a high cognitive load, on the other hand, can be assumed to benefit from a neural environment capable of adapting to variations in input and

associations between different kinds of information and therefore show a positive correlation with cortex thickness. Previous studies have shown that musicians with absolute pitch have thinner cortex in RIFG pars opercularis, anterior middle frontal gyrus, and premotor cortex (Bermudez, Lerch, Evans, & Zatorre, 2009). In addition, it has been seen that event-related potentials related to basic auditory processing negatively correlate with cortical thickness in the right supratemporal plane (Liem, Zaehle, Burkhard, Jancke, & Meyer, 2012). Conversely, a positive correlation has been found between cortical thickness in left planum temporale and pars opercularis of LIFG and speed in associating language tones with grammatical endings (Schremm et al., 2018), a task which can be assumed to induce a high cognitive load. Likewise, cortical thickness in various cortical areas has been seen to correlate positively with general intelligence measures (Menary et al., 2013). The LLAMA F test requires participants to infer how the morphosyntactic units and descriptive figures interact. In the pitch discrimination task, however, participants are instructed to determine how two elements of the same kind (tones) relate to each other. Degree of element interactivity is a driver of cognitive load (Paas, Renkl, & Sweller, 2003; Sweller, 1988). Hence, the LLAMA F test, involving interaction of several elements, can be assumed to induce a higher cognitive load than the pitch discrimination test. A possible generalization based on both previous studies and results presented in this study could be that skills involving a higher cognitive load require a richer representation of the problem to solve, thus necessitating larger and more intricate networks of cells, leading to a thicker cortex, while skills associated with relatively low cognitive load benefit more from fast processing in a relatively thinner, more specialized cortical structure.

By learning more about which neural substrates contribute to language learning aptitude, research on enhancing learning or the impact of neurodevelopmental disorders from a neurophysiological perspective is made possible. The correlations with LLAMA F scores in anterior Broca's area could provide such information but the lack of correlations for the LLAMA B and D scores could indicate that vocabulary or implicit

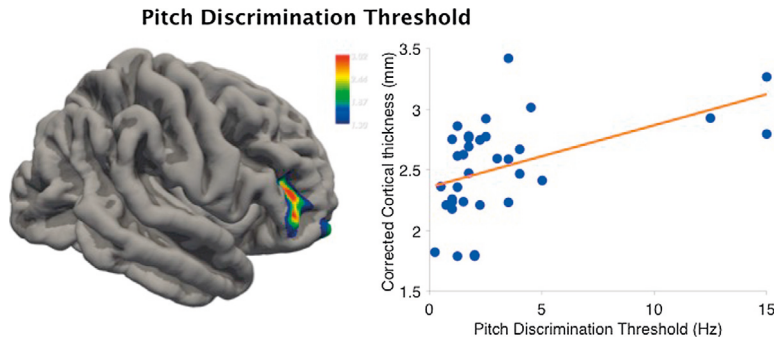


Fig. 2. Left: Significant clusters showing positive correlation between pitch discrimination threshold and cortical thickness given age and eICV. Color bar indicating Z statistic. Right: Adjusted response plots from max voxel from right inferior frontal gyrus.

phonetic memory can be taught in many different ways whereas grammatical meaning inferring is a skill that more heavily relies on specific cortical areas. Speculating, this could indicate that more resources might have to be concentrated on one way of learning grammatical meaning inferring while students probably acquire vocabulary and speech pattern recognition best overall if they are offered a plentitude of learning methods.

In summary, our results indicate that the cortical thickness of LIFGpt and LMFG has importance for proficiency in grammatical meaning inferring skills and the cortical thickness of the RIFG for accurate pitch direction comparison. The results are in line with findings showing a tendency for complex skills associated with a high cognitive load to be aided by a thicker cortex, whereas skills involving a relatively low cognitive load instead benefit from a thinner, more specialized cortex. These findings contribute to extending our understanding of the relationship between cortical architecture and aptitude for language learning.

4. Method

4.1. Subjects

44 right-handed healthy volunteers (20–32 years of age) were recruited. 24 (12 female, 12 male) of the subjects were native speakers of Swedish and 20 (9 female, 11 male) were native speakers of German. Handedness was assessed through the Edinburgh Handedness Inventory (Oldfield, 1971).

4.2. Audiometry

Fixed-frequency Békésy audiometry was used to estimate subjects' pure tone hearing thresholds for frequencies 250, 500, 1000, 2000, 4000, and 8000 Hz using a GN Otometrics Astera audiometer and a pair of circumaural sound attenuating Sennheiser HDA 200 earphones. The stimuli consisted of pulsed pure tones gated on 250 ms with 250 ms silent interval between presentations and a 2 dB intensity rate change per second was used. Six reversals were recorded per frequency and the threshold at each test frequency was calculated as the mean of these six reversals. The calibration of this set-up was made using a Brüel and Kjaer 2231 sound level meter with a 4134 microphone in a 4153 artificial ear according to IEC 60318-2 and ISO 389-8 (IEC, 1998; ISO, 2004). All subjects had normal hearing defined as pure-tone hearing thresholds < 20 dB hearing level (ISO, 2004) for all presented frequencies.

4.3. Magnetic resonance imaging scans

MRI was performed with an actively shielded 7T scanner (Achieva, Philips, Best, Netherlands) equipped with a dual transmit head coil and a 32 channel receive phased-array head coil (Nova Medical, Wilmington, MA). In order to address inherent inhomogeneities due to the high field strength, one T1 and one proton density (PD) weighted volume were acquired using 3D magnetization-prepared rapid gradient echo (MPRAGE) sequences with field of view = $200 \times 252 \times 190$ mm³, repetition time = 5 ms, echo time = 2 ms, flip angle = 6°, inversion pulse delay = 1200 ms, and repetition time = 6 ms, echo time = 2.5 ms, flip angle = 7°, respectively. This approach is based on the concept of MP2RAGE (Marques et al., 2010). Dielectric pads were used to reduce the effect of B1-field inhomogeneities (Teeuwisse, Brink, & Webb, 2012). Data from 7 (1 female, 6 male) Swedish and 3 (1 male, 2 female) German native speakers, were discarded from further analysis due to image artifacts.

4.4. Image analysis

For each subject, the PD-weighted volume was registered to the T1-weighted volume. Subsequently, the T1-weighted volume was divided by the PD-weighted and improved nonparametric nonuniform intensity normalization bias field correction was applied (version 2.1.0.post685-g86e09) (Tustison et al., 2010).

Cortical reconstruction and volumetric segmentation was performed with the Freesurfer image analysis suite v6.0. The technical details of these procedures have been described elsewhere (Dale, Fischl, & Sereno, 1999; Fischl & Dale, 2000; Fischl, Liu, & Dale, 2001; Fischl et al., 2002; Fischl, Salat, et al., 2004; Fischl, Sereno, Tootell, & Dale, 1999; Fischl, van der Kouwe, et al., 2004; Han et al., 2006; Reuter, Rosas, & Fischl, 2010; Reuter, Schmansky, Rosas, & Fischl, 2012; Ségonne et al., 2004). Cortical thickness data from each subject was transformed into Freesurfer common space, fsaverage, before a 10 mm full width at half maximum Gaussian spatial smoothing kernel was applied.

4.5. Pitch perception & language aptitude tests

The LLAMA tests were developed with the ambition of assessing the aptitude of learning a foreign language, independent of the first language of the test-taker (Meara, 2005). Of the four available tests, participants in the present study completed LLAMA B, D, and F. The task in LLAMA B consists of memorizing words (letter strings) associated with pictures. Phonetic memory is assessed in LLAMA D by presenting the test-participants with 10 spoken foreign words after which they listen to

another set of words and report which words are familiar and which are not. LLAMA F focuses on the ability to infer the grammatical meaning of morphosyntactic units on the basis of sentences describing events illustrated by pictures. 20 pictures of bipedal figures of different colors at no or different relative positions to a rectangular block are presented in a 5 min training phase during which the test takers are allowed to take notes. After training, novel pictures are presented along with two sentences, one correct and one incorrect, and the participant is asked to choose which one he or she believes to be correct in the target language.

Since proficiency in pitch discrimination is not evaluated in the LLAMA tests, the present study chose to assess it using an in-house developed computer-based test in which subjects listened to two tones in succession that differed in pitch. The first tone was always 550 Hz. The difference in pitch ranged from 30 to 0.25 Hz. Subjects were asked to judge whether the second tone was higher or lower in pitch than the first and were informed on the validity of their answer. If the number of correct answers was more than three above the number of false ones at a given point in the test, the difference in pitch was reduced, while if the number of false answers was above the number of correct ones, the difference in pitch was increased. The score used for analysis was the minimal pitch difference reached after 100 repetitions. This score was taken as a measure of the minimum pitch difference a subject needed to discriminate pitch change direction: the lower the score, the greater the proficiency in pitch discrimination.

4.6. Experimental design and statistical analysis

A separate correlation analysis was performed for each behavioral measure. Cortical thickness is known to decrease with age (Thambisetty et al., 2010). Moreover, brain size is positively correlated with cortical thickness (Im et al., 2008). To take into account these effects, age (in years), and ICV, given from FreeSurfer output, were included as covariates. Clusters in which cortical thickness correlated with grammar aptitude scores or pitch discrimination threshold underwent comparison to cluster size limits derived from Monte Carlo simulation based multiple comparison correction using a cluster-wise threshold of $p < 0.05$ (Hagler, Saygin, & Sereno, 2006). Following recommendations for reporting fMRI results, the statistics for the maximum vertex in the clusters were reported (Poldrack et al., 2008).

5. Statement of significance

We found that cortical thickness in Broca's area and its right homologue are associated with grammar learning aptitude and skill in pitch discrimination, both of importance for language learning. For the first time, we show that cortical thickness can be an important factor underlying language learning aptitude.

Author contributions

Novén and Schremm designed and performed the experiments. Novén analysed the data and wrote the manuscript. Roll supervised the experiments and analysis. Nilsson implemented the MRI sequences. All authors discussed the results and implications and commented on the manuscript at all stages.

Conflict of interest

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

Funding

This work was supported by Knut and Alice Wallenberg Foundation

(grant number 2014.0139), Marcus and Amalia Wallenberg Foundation (grant number 2014.0039), and the Swedish Research Council (grant number 2011-2284).

Acknowledgements

Lund University Bioimaging Center (LBIC), Lund University, Sweden, is gratefully acknowledged for providing experimental resources.

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>.
- Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex*, 19(7), 1583–1596.
- Carroll, J. B., & Sapon, S. (1959). *Modern language aptitude test*. San Antonio, TX: Psychological Corporation.
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. *NeuroImage*, 9(2), 179–194.
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, 97(20), 11050–11055.
- Fischl, B., Liu, A., & Dale, A. M. (2001). Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Transactions on Medical Imaging*, 20(1), 70–80.
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation. *Neuron*, 33(3), 341–355.
- Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., et al. (2004). Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23, 569–584.
- Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. *NeuroImage*, 9(2), 195–207.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., et al. (2004). Automatically parcellating the human cerebral cortex. *Cereb Cortex*, 14(1), 11–22.
- 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.
- Forsberg Lundell, F., & Sandgren, M. (2013). High-level proficiency in late L2 acquisition. Relationships between collocational production, language aptitude and personality. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 231–256). Amsterdam: John Benjamins.
- Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, 89(2), 267–276.
- Goel, V., Gold, B., Kapur, S., & Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *Neuroreport*, 8(5), 1305–1310.
- 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.
- Gosselke Berthelsen, S., Horne, M., Bränström, K. J., Shtyrov, Y., & Roll, M. (2018). Neural processing of morphosyntactic tonal cues in second-language learners. *Journal of Neurolinguistics*, 45(Supplement C), 60–78.
- Goucha, T., & Friederici, A. D. (2015). The language skeleton after dissecting meaning: A functional segregation within Broca's Area. *NeuroImage*, 114, 294–302.
- Granena, G. (2012). Age differences and cognitive aptitudes for implicit and explicit learning in ultimate second language attainment (Doctoral. University of Maryland, Unpublished).
- 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>.
- 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.
- Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2004). Structural brain variation and general intelligence. *NeuroImage*, 23(1), 425–433.
- Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., et al. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage*, 32(1), 180–194.
- Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: When less is better than more. *Journal of Neuroscience*, 27(47), 13028–13032.
- IEC (1998). IEC 60318-2. Electroacoustics – Simulators of human head and ear – Part 2: An interim acoustic coupler for the calibration of audiometric earphones in the extended high-frequency range. International Electrotechnical Commission, Geneva.
- 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. *Cereb Cortex*, 18(9), 2181–2191.
- ISO (2004). ISO 389-8. Acoustics: Reference zero for the calibration of audiometric equipment. Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones. International Organization for Standardization

- 389–8.
- Karama, S., Ad-Dab'bagh, Y., Haier, R. J., Deary, I. J., Lyttelton, O. C., Lepage, C., & Brain Development Cooperative Group (2009). Positive association between cognitive ability and cortical thickness in a representative US sample of healthy 6 to 18 year-olds. *Intelligence*, *37*(2), 145–155.
- Klein, D., Mok, K., Chen, J.-K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, *131*, 20–24.
- Lee, H., Devlin, J. T., Shakeshaft, C., Stewart, L. H., Brennan, A., Glensman, J., et al. (2007). Anatomical traces of vocabulary acquisition in the adolescent brain. *Journal of Neuroscience*, *27*(5), 1184–1189. <https://doi.org/10.1523/jneurosci.4442-06.2007>.
- Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*, *58*(Supplement C), 301–324.
- Liem, F., Zaehle, T., Burkhard, A., Jancke, L., & Meyer, M. (2012). Cortical thickness of supratemporal plane predicts auditory N1 amplitude. *Neuroreport*, *23*(17), 1026–1030.
- Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., & Gruetter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage*, *49*(2), 1271–1281.
- Marslen-Wilson, W. D., & Tyler, K. D. (2007). Morphology, language and the brain: The decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society B*, *362*, 823–836.
- Meara, P. (2005). *LLAMA language aptitude tests: The manual*. Swansea, UK: Lognostics.
- Menary, K., Collins, P. F., Porter, J. N., Muetzel, R., Olson, E. A., Kumar, V., et al. (2013). Associations between cortical thickness and general intelligence in children, adolescents and young adults. *Intelligence*, *41*(5), 597–606.
- Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., et al. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, *63*(1), 240–244.
- Nauchi, A., & Sakai, K. L. (2009). Greater leftward lateralization of the inferior frontal gyrus in second language learners with higher syntactic abilities. *Human Brain Mapping*, *30*(11), 3625–3635.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4.
- Perrachione, T. K., Ghosh, S. S., Ostrovskaya, I., Gabrieli, J. D. E., & Kovelman, I. (2017). Phonological working memory for words and nonwords in cerebral cortex. *Journal of Speech, Language, and Hearing Research*, *60*(7), 1959–1979. https://doi.org/10.1044/2017_JSLHR-L15-0446.
- Peterson, K. M., & Hagoort, P. (2012). The neurobiology of syntax: Beyond string sets. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *367*(1598), 1971–1983.
- Poldrack, R. A., Fletcher, P. C., Henson, R. N., Worsley, K. J., Brett, M., & Nichols, T. E. (2008). Guidelines for reporting an fMRI study. *NeuroImage*, *40*(2), 409–414. <https://doi.org/10.1016/j.neuroimage.2007.11.048>.
- Porter, J. N., Collins, P. F., Muetzel, R. L., Lim, K. O., & Luciana, M. (2011). Associations between cortical thickness and verbal fluency in childhood, adolescence, and young adulthood. *NeuroImage*, *55*(4), 1865–1877.
- Posedel, J., Emery, L., Souza, B., & Fountain, C. (2011). Pitch perception, working memory, and second-language phonological production. *Psychology of Music*, *40*(4), 508–517.
- Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., et al. (2011). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: Functional neuro-imaging and brain morphology. *Frontiers in Psychology*, *2*, 271.
- Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, *53*(4), 1181–1196.
- Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, *61*(4), 1402–1418.
- Roll, M., Söderström, P., Mannfolk, P., Shtyrov, Y., Johansson, M., van Westen, D., et al. (2015). Word tones cueing morphosyntactic structure: Neuroanatomical substrates and activation time-course assessed by EEG and fMRI. *Brain and Language*, *150*(Supplement C), 14–21.
- Schremm, A., Novén, M., Horne, M., Söderström, P., van Westen, D., & Roll, M. (2018). Cortical thickness of planum temporale and pars opercularis in native language tone processing. *Brain and Language*, *176*, 42–47.
- Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., et al. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, *22*(3), 1060–1075.
- Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency. *Psychological Science*, *17*(8), 675–681.
- Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004). Longitudinal mapping of cortical thickness and brain growth in normal children. *The Journal of Neuroscience*, *24*(38), 8223.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.
- Söderström, P., Horne, M., & Roll, M. (2017). Stem tones pre-activate suffixes in the brain. *Journal of Psycholinguistic Research*, *46*(2), 271–280.
- 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.
- Thambisetty, M., Wan, J., Carass, A., An, Y., Prince, J. L., & Resnick, S. M. (2010). Longitudinal changes in cortical thickness associated with normal aging. *NeuroImage*, *52*(4), 1215–1223.
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320.
- Wang, W. S. Y. (1973). The Chinese language. *Scientific American*, *228*(2), 50–63.
- 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.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46.

Study IV



Cortical and white matter correlates of language learning aptitudes

Mikael Novén^a, Hampus Olsson^b, Gunther Helms^b, Merle Horne^a, Markus Nilsson^c & Mikael Roll^a

^a Department of Linguistics and Phonetics, Lund University, 221 00, Sweden

^b Department of Medical Radiation Physics, Clinical Sciences Lund, Lund University, Lund, Sweden

^c Department of Clinical Sciences Lund, Radiology, Lund University, 221 00, Sweden

Acknowledgements

This work was supported by Knut and Alice Wallenberg Foundation (Grant No. 2014.0139) and the Swedish Research Council (Grant No. 2018.00632). Lund University Bioimaging Center (LBIC), Lund University, Sweden, is gratefully acknowledged for providing experimental resources.

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 inferencing 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-millimetre 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 on negative correlations between scores for phonetic memory and axial kurtosis in left arcuate fasciculus and superior longitudinal fasciculus III, which are tracts connecting cortical areas important for phonological working memory.

Keywords

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

Introduction

The study of what makes a good language learner has been a growing research field since at least the 1950's (Carroll & Sapon, 1959). The need and benefit for people to learn new languages far into adulthood has only grown with globalisation, 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 be useful for studying the neural correlates of 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). Furthermore, Golestani (2012) argues that research into the brain structural correlates of innate aptitude might be worthwhile since evidence from expert

phoneticians suggest a difference in the structure of the Heschl's gyrus (HG) relative to non-experts likely present prior to any training (Golestani, Price, & Scott, 2011). This has motivated research into the associations between language learning aptitude and cortical morphology (Novén, Schremm, Nilsson, Horne, & Roll, 2019; Rodriguez, Archila-Suerte, Vaughn, Chiarello, & Hernandez, 2018; Turker, Reiterer, Schneider, & Seither-Preisler, 2019; Turker, Reiterer, Seither-Preisler, & Schneider, 2017; P. C. M. Wong et al., 2008) as well as white matter microstructure (Flöel, de Vries, Scholz, Breitenstein, & Johansen-Berg, 2009; Mårtensson et al., 2020; Qi, Han, Garel, San Chen, & Gabrieli, 2015; Xiang et al., 2012). These studies have found effects of the grey matter volume or shape of HG in learning Mandarin word tones (P. C. M. Wong et al., 2008), performance on a speech imitation task including novel phonological contrasts (Turker et al., 2017), and general language learning aptitude (Turker et al., 2019). In addition, cortical thickness of anterior insula correlates with aptitude for learning to discriminate new speech sounds in bilinguals but not monolinguals (Rodriguez et al., 2018). Furthermore, cortical thickness of Brodmann area (BA) 45 in the left inferior frontal gyrus (IFG) has been found to correlate with grammatical inferencing ability (Novén et al., 2019). Studies including measures of white matter microstructure have instead found associations between artificial grammar learning ability and the structure of connections to the left IFG (Flöel et al., 2009) and structure of right-hemispheric white matter and successful learning of Mandarin (Qi et al., 2015). Furthermore, white matter microstructure in frontal networks predicts language learning proficiency in conscript interpreters (Mårtensson et al., 2020). However, previous studies have been limited in several aspects. 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; Rodriguez et al., 2018; Turker et al., 2017; P. C. M. Wong et al., 2008). These approaches however, are not specifically related to language learning aptitude theory, which strives towards understanding what cognitive processes are involved in making a good language learner (Wen, Biedroń, & Skehan, 2016). In fact, only three studies we know of, apart from our previous work (Novén et al., 2019), have examined brain structural correlates of behavioural measures rooted in language learning aptitude theory (Turker et al., 2019; Xiang et al., 2012). These studies are described in detail below. With regards to parameters of brain structure that are studied, some have only investigated cortical morphology in predefined regions of interest (ROIs) (Rodriguez et al., 2018; Turker et al., 2019; Turker et al., 2017; P. C. M. Wong et al., 2008) while the others have been limited in spatial resolution (Novén et al., 2019). Moreover, 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). By 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, i.e. the thickness and surface area of the cortex, at sub-millimetre resolution. Further, DKI is used to investigate 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 allows for measuring anatomical correlates of language learning aptitude in more detail compared to previous studies.

LLAMA test battery and brain structure

Language learning aptitude can be assessed using the first language-independent, computer-based LLAMA test battery (Meara, 2005) consisting of 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 LLAMA test battery has been used previously to investigate correlates with brain anatomy. Grey matter volume in 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 BA 45 in each hemisphere (Xiang et al., 2012). Moreover, vocabulary learning aptitude has been shown to correlate with a difference between FA in left and right tracts between BA47 and the parietal lobe (Xiang et al., 2012). In our previous work, we could show 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 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 found 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 studied 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), a method which runs the risk of including non-anatomical white matter tracts (Schilling et al., 2019).

White matter tract segmentation

There are four anatomical white matter tracts that are known to be important for language processing and that we believe could be relevant correlates for language learning aptitudes. These are 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 inferior frontal gyrus (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; F. C. K. 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 (F. C. K. 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 e.g. Pierpaoli, Jezzard, Basser, Barnett, and Di Chiro (1996) and Jensen and Helpert (2010) for DTI and DKI, respectively. Put simply, DTI yields the mean diffusivity (MD) which is the mean apparent diffusivity across all diffusion encoding angles while axial and radial diffusivity (AD and RD) are the diffusivities along and perpendicular to the direction of greatest diffusivity. Fractional anisotropy is a measure of 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). If 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 standardised and normalised 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. 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 non-anatomical, i.e. 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, i.e. 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.

The present study

The aim of the present study was to investigate 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 (F. C. K. Wong et al., 2011; Xiang et al., 2012), as a basis for cortical morphometry. Furthermore, utilising DKI allows for calculation of more detailed tissue microstructure parameters of relevant white matter tracts as compared to standard DTI.

Method

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; the oldest participant was 27 and the youngest 20 years old. No participant had any history of psychiatric disorders. All participants were right-handed as defined as a minimum of +25 in Edinburgh handedness index (Oldfield, 1971) and had vision that was normal or corrected to normal. To ensure normal hearing (important for the phonetic memory subtest), a minimal hearing threshold of <20dB for pure tones of 250, 500, 1000, 2000, 4000, and 8000 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 characterised with respect to 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 detail in Mårtensson and Lövdén (2011). Participants were given 10 minutes to complete 18 matrices by selecting the missing 9th pattern based on eight given patterns. Scores were equal to the number of correctly selected 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.

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 twenty cartoon objects and were given the written name of each object when clicking on them with the computer mouse. Each participant was given two minutes 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 closely listen to a short set of spoken foreign words. Immediately afterwards, 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 twenty 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 five-minute 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) sentences 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.

MRI acquisition

MRI was performed with an actively shielded 7T scanner (Achieva, Philips, Best, Netherlands) equipped with a dual transmit head coil and a 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 magnetisation-prepared rapid gradient echo (MPRAGE) sequence with repetition time (TR) = 8ms, echo time (TE) = 1.97 ms, flip angle = 8°, inversion time (TI) = 1200 ms, 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 normalisation, a proton density (PD)-weighted gradient echo sequence at identical TR and TE = 1.97 but lower flip angle = 2°, resolution ($1.6 \times 1.6 \times 1.6 \text{ mm}^3$) with a scan time of 54 s was also acquired.

Two diffusion-weighted sequences were acquired, one with a higher number of encoding directions for better tractography and one with a higher number of shells for diffusion kurtosis tensor estimations. The first 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 $b = [0 \ 2] \text{ ms}/\mu\text{m}^2$ 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 second 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, $b = [0 \ 0.1 \ 0.7 \ 1.4 \ 2] \text{ ms}/\mu\text{m}^2$ and $[1 \ 6 \ 6 \ 12 \ 16]$ encoding directions giving a scan time of 4:40 min. The ordering of the encoding strengths were randomised in order to temporally spread the energy consumption and avoid temporal confounds (Hutter et al., 2018; Vos et al., 2017).

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 further mitigate B1 inhomogeneities and cancel weighting by PD and effective transverse ($T2^*$) relaxation. This approach is based on the concept described in Van de Moortele et al. (2009). Image volumes were brain extracted using the brain extraction tool 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 normalised and brain-extracted T1-weighted image volumes then underwent nonparametric non-uniform bias field correction, to again reduce 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, Segonne, 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 except for the skull stripping (performed separately as mentioned above) with the `-hires` flag (Zaretskaya et al., 2018). The processing pipeline

included motion-correction, intensity normalisation, 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.

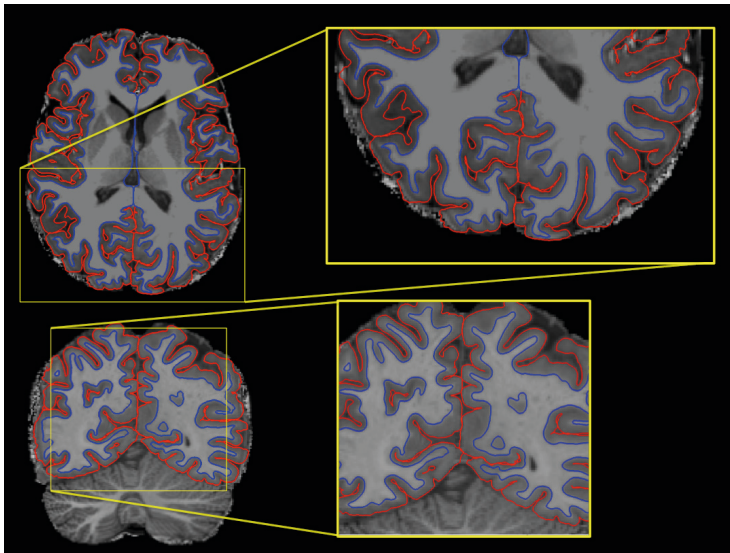


Figure 1: Example of the resulting FreeSurfer surface reconstructions. Pial surfaces are shown in red and white surfaces in blue.

DWI processing

Two diffusion-weighted image volumes were acquired. One single-shelled with a higher directional resolution, henceforth called the tractography volume, and one multi-shelled, hereafter the DKI volume, to permit for DKI estimation.

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 as well as the DKI volume were 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 volumes to MNI space using an FA-template included in FSL. From the diffusion-weighted MRI data, it is possible to segment anatomical connections between cortical areas, i.e. the white matter fibre tracts. This could be done manually using sets of inclusion and exclusion ROIs but this runs the risk of introducing a great deal of subjectivity in the placement of the

ROIs and inclusion of many non-anatomical white matter tracts. Therefore, white matter bundle segmentation was performed on the pre-processed tractography volume using the TractSeg tool openly available at <https://github.com/MIC-DKFZ/TractSeg> (Wasserthal et al., 2019; Wasserthal et al., 2018a, 2018b). 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 fibre 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 motion-corrected DKI volume underwent correction for noise, Gibbs ringing, Rician biases, and signal outliers as well as diffusion tensor and kurtosis parameter estimation using the freely available DESIGNER tool (<https://github.com/NYU-DiffusionMRI/DESIGNER>) (Ades-Aron et al., 2018). Prior to parameter estimations, the corrected DKI volumes were smoothed using a 2.4 mm FWHM Gaussian kernel. The mean parameter values of each tract were then extracted using the Tractometry command in TractSeg. However, the IFOF segmentation mapped poorly to the DKI volume due to too much signal loss in the inferior parts of the brain in the DKI volume. Parameters from IFOF were only obtainable from 10 of the subjects. Therefore, IFOF was not included in the analyses.

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.

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. Age was a covariate of no interest in all cortical and diffusion parameter analyses. 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 (6000 permutations)-based multiple comparison correction (Hagler, Saygin, & Sereno, 2006) using a cluster-forming threshold of $p < 0.005$ or $p < 0.001$ for thickness and surface area, respectively, in accordance with suggestions in Greve and Fischl (2018). Results were projected onto the pial surface of `FsAverage` for display purposes. All other statistical analyses were performed using R (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 supplementary materials. The significance threshold from diffusion parameter

correlation tests was Bonferroni corrected for multiple comparisons across the four analysed white matter tracts, yielding a significance threshold of $p=0.05/4=0.0125$.

Results

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 ρ 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, i.e. that they captured separate components of language learning aptitude.

Table 1: Spearman's rho and uncorrected p-values for comparisons between each background measure. * $p<0.05$; ** $p<0.01$

	LLA MA B	LLAMA D	LLAMA F	OSPAN	Raven	Musicality general sophistication	Number of learned languages
LLAMA B (Vocabulary)	1	$\rho = -0.15$ $p = 0.27$	$\rho = -0.11$ $p = 0.40$	$\rho = 0.25$ $p = 0.062$	$\rho = 0.01$ $p = 0.94$	$\rho = 0.12$ $p = 0.38$	$\rho = -0.13$ $p = 0.34$
LLAMA D (Phonetics)		1	$\rho = -0.09$ $p = 0.51$	$\rho = 0.09$ $p = 0.51$	$\rho = 0.12$ $p = 0.38$	$\rho = -0.08$ $p = 0.54$	$\rho = 0.12$ $p = 0.38$
LLAMA F (Grammar)			1	$\rho = -0.11$ $p = 0.41$	$\rho = 0.29$ $p = 0.032^*$	$\rho = 0.04$ $p = 0.78$	$\rho = 0.19$ $p = 0.15$
OSPAN				1	$\rho = 0.17$ $p = 0.22$	$\rho = 0.19$ $p = 0.16$	$\rho = 0.28$ $p = 0.034^*$
Raven					1	$\rho = 0.00$ $p = 0.98$	$\rho = 0.36$ $p = 0.0063^{**}$
Musicality general sophistication						1	$\rho = 0.06$ $p = 0.63$
Number of learned languages							1

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 to 0.93 $\mu\text{m}^2/\text{ms}$, 0.31 to 0.70 $\mu\text{m}^2/\text{ms}$, 0.41 to 0.83, 0.81 to 1.32, and 1.02 to 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 III, values for UF were unfeasible for white matter tracts. Hence, values from UF were excluded from further analyses.

Table 2: Mean diffusion parameter values for each studied white matter tract. 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.

Diffusion parameter	MD	FA	MK	AD	RD	AK	RK
Left AF	0.89 ± 0.03	0.36 ± 0.02	1.15 ± 0.04	1.23 ± 0.03	0.72 ± 0.04	0.93 ± 0.03	1.43 ± 0.08
Right AF	0.88 ± 0.03	0.36 ± 0.03	1.12 ± 0.05	1.21 ± 0.03	0.71 ± 0.04	0.94 ± 0.04	1.35 ± 0.10
Left SLF III	0.88 ± 0.03	0.38 ± 0.02	1.14 ± 0.05	1.23 ± 0.03	0.71 ± 0.04	0.91 ± 0.04	1.40 ± 0.08
Right SLF III	0.89 ± 0.03	0.36 ± 0.02	1.13 ± 0.05	1.22 ± 0.03	0.72 ± 0.03	0.94 ± 0.04	1.38 ± 0.11
Left UF	1.26 ± 0.23	0.22 ± 0.04	0.72 ± 0.11	1.57 ± 0.29	1.09 ± 0.20	0.76 ± 0.12	0.67 ± 0.11
Right UF	1.12 ± 0.12	0.23 ± 0.03	0.68 ± 0.06	1.39 ± 0.15	0.97 ± 0.09	0.69 ± 0.06	0.64 ± 0.08

LLAMA B

Cortical surface area in the left posterior inferior precuneus correlated with LLAMA B scores, as shown in table 3 and figure 2. No correlations were found between any diffusion parameter in any tract and the LLAMA B score (see supplementary details for more information).

Table 3: Statistics for cluster in left posterior inferior precuneus in which cortical surface area correlated with LLAMA B score.

Cortical areas	Cluster size (mm^2)	Pearson's r	Peak Z	Peak Z coordinates (MNI x,y,z)	Cluster-wise p
L. Precuneus	325	0.485	4.88	-8, -55, 17	0.00840

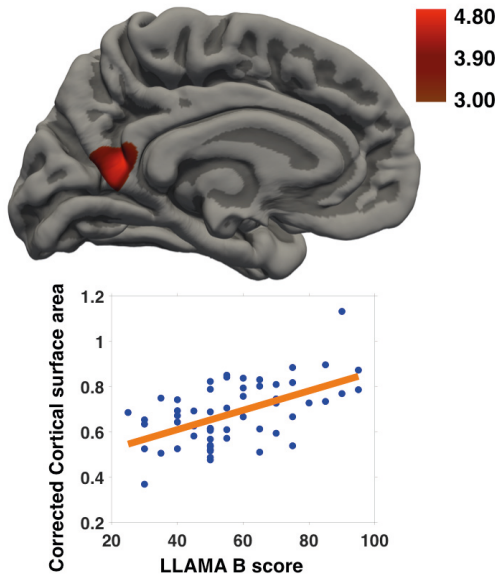


Figure 2: Vocabulary learning aptitude score (LLAMA B) correlated with cortical surface area in a left posterior medial cluster covering the left inferior posterior precuneus, given eICV and age as covariates.

LLAMA D

Mean AK along the left AF and left SLF III correlated with phonetic memory (LLAMA D) score ($r=-0.523$, $t(52)=-4.40$, $p=5.92e-5$) and ($r=-0.458$, $t(52)=-3.68$, $p=5.70e-4$), respectively (figure 3). 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 supplementary information for details).

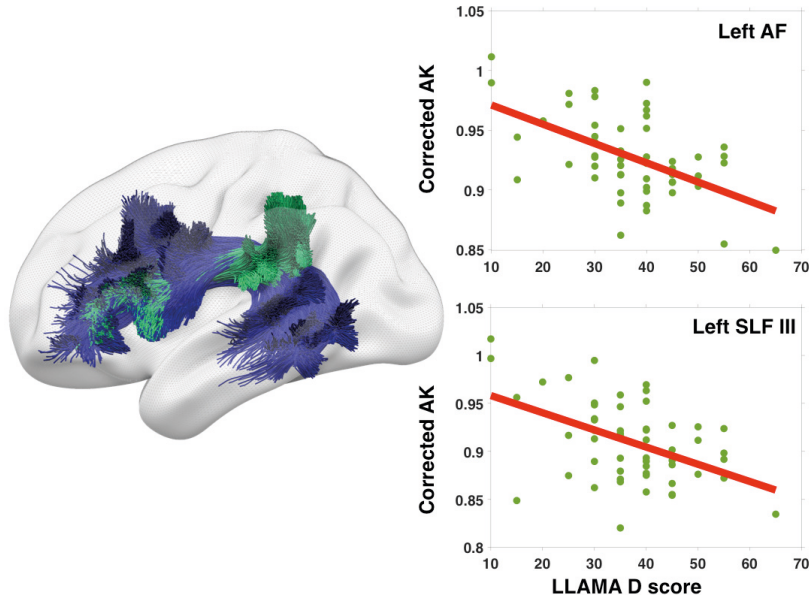


Figure 3: Phonetic memory aptitude (LLAMA D) correlated with axial kurtosis (AK) in left arcuate fasciculus (AF) and left superior longitudinal fasciculus part III. AF (blue) and SLF III (green) are shown in the image to the left, created using SurfIce (freely available at <https://github.com/neurolabusc/surf-ice>).

LLAMA F

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

Discussion

This study set out to investigate cortical morphometric and white matter fibre tract diffusion parameter correlates of language learning aptitude. By finding neural foundations for language learning abilities, we learn 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 correlates with AK in left AF and SLF III. A negative correlation with AK can be thought of as phonetic memory benefitting from a more coherent and homogeneous nerve fibre tract. Compared with previous studies 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 in our supplementary materials) 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) investigated children between 10 and 16 years of age whereas our participants ranged between 20 to 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, Ségonne, 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 right HG and LLAMA subtest scores, reported p-values range from 0.003 to 0.028. It should be noted that we found a non-significant correlation between cortical thickness in the right HG and LLAMA F score ($r=0.34$, uncorrected $p=0.013$), see supplementary 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).

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 known to be involved in integration between visuo-spatial inputs and declarative memory (Cavanna & Trimble, 2006). The inferior posterior part, specifically, has been shown to be part of the hippocampal-cortical memory system (Kahn, Andrews-Hanna, Vincent, Snyder, & Buckner, 2008; Margulies et al., 2009; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008; Vincent et al., 2006). Left precuneus has a different cortical developmental trajectory in bilingual children and young adults, linked to increased demands on lexical knowledge and, to some extent, knowledge of grammars and phonological systems (Pliatsikas et al., 2020). Left precuneus has also been found to be activated by remembered as compared to forgotten words one week after practicing in a vocabulary learning experiment (naïve learners of Swahili) (van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013). In accordance with said findings, 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 found in (Xiang et al., 2012). This could be due to the difference in tract segmentation approach. While Xiang and colleagues (2012) used 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.

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 as well as the left SLF III. The AF connects the IFG with the middle frontal gyrus, the posterior superior temporal gyrus as well as the temporal occipital transition region, while the adjacent SLF III connects the supramarginal gyrus with the prefrontal and ventral premotor cortices. AK is an index of tissue complexity along the principal diffusion direction, i.e. along the

fibres in an ideal white-matter voxel. Higher AK could thus be due to the presence of non-axonal cell membranes, e.g. glial cells, astrocytes, and oligodendrocytes (Hui, Cheung, Qi, & Wu, 2008) or tortuosity of the axons (Fieremans, Jensen, & Helpert, 2011). An intuitive interpretation of the results would consequently be that more coherent and more homogeneous left AF and SLF III are beneficial for phonetic memory. Both left AF and SLF III are part of the dorsal language processing stream and involved in repetition of especially pseudowords (Hickok & Poeppel, 2004; Saur et al., 2008).

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 capacities. Moreover, learners that are faster at perceiving novel speech sound contrasts have more white matter in (especially left) parietal regions (Golestani, Paus, & Zatorre, 2002). Success in learning of Mandarin Chinese words from listening to a short film has been seen to be associated with greater functional connectivity between 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.

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 key 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.

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 find no correlations at all with cortical thickness in the present study; a table with mean values of cortical thickness and surface area in a number of language-relevant cortical areas are provided as supplementary material. 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% as compared to 31.5% in this study. However, the LLAMA F test scores did not differ significantly between this study ($M=63.5$, $SD=19.7$) and Novén et al. (2019) ($M=67.9$, $SD=22.8$) ($t(62.8)=-0.932$, $p=0.355$, Welch two sample t-test). 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 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 with the exception of a moderate link between fluid intelligence and grammatical inferencing score.

Limitations and future directions

We identified four primary types of limitations in the present study. First, we did the imaging at 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 artefacts 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 supplementary material. 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 artefacts 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/fibre 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/aptitude for what the component captures.

The fourth limitation is that this study could potentially be limited in statistical power. A reasonable estimate (based on the studies cited in this article) of the correlation coefficients we can expect in correlations between behavioural measures and diffusion parameters or cortical morphometrical measures range from $r=0.35$ to 0.56 . Given a significance threshold of 0.05 , we thus need between 23 and 62 participants to detect such correlation at a power of 0.8 . This is assuming only one ROI but in neuroimaging studies, it is normal to include at least the contralateral homologue of the ROI. If the significance threshold is Bonferroni corrected in accordance 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 $r=0.46$ to 0.66 but have to be corrected for at least 58 ROIs (see reasoning above), and thus require 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.

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 axial kurtosis 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 fibre tracts correlate with talent for learning languages. In a wider perspective, this also means that difference in aptitudes and talents could possibly, at least partly, be due to differences in cortical morphometry or white matter microstructure.

References

- Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, 30(4), 481-509. doi: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. doi: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. doi: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. doi: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. *Ann Neurol*, 57(1), 8-16. doi: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. doi:http://dx.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. doi:10.1093/brain/awl004
- 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. doi: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. doi: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. doi:10.1016/j.neuroimage.2006.01.021
- Desikan, R. S., Ségonne, 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. doi:http://dx.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. doi: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. doi:10.1016/j.neuroimage.2011.10.065
- Fieremans, E., Jensen, J. H., & Helpert, J. A. (2011). White matter characterization with diffusional kurtosis imaging. *NeuroImage*, *58*(1), 177-188. doi: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. doi:https://doi.org/10.1006/ning.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. doi: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. doi: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. doi: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. doi:10.1177/1367006912456585
- Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical Correlates of Learning Novel Speech Sounds. *Neuron*, *35*(5), 997-1010. doi: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. doi: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. doi:10.1177/0267658312461497
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63-72. doi: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. doi: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. doi: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. doi:http://dx.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. doi:<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. *Cereb Cortex*, 18(9), 2181-2191. doi: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. doi:10.1016/s1053-8119(02)91132-8
- Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med Image Anal*, 5(2), 143-156. doi: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 Biomed*, 23(7), 698-710. doi:10.1002/nbm.1518
- Kahn, I., Andrews-Hanna, J. R., Vincent, J. L., Snyder, A. Z., & Buckner, R. L. (2008). Distinct cortical anatomy linked to subregions of the medial temporal lobe revealed by intrinsic functional connectivity. *Journal of Neurophysiology*, 100(1), 129-139. doi:10.1152/jn.00077.2008
- 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. doi:10.1093/applin/amu054
- 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. doi:10.1002/jmri.23857
- 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. doi:10.1093/cercor/bhh186
- Margulies, D. S., Vincent, J. L., Kelly, C., Lohmann, G., Uddin, L. Q., Biswal, B. B., . . . Petrides, M. (2009). Precuneus shares intrinsic functional architecture in humans and monkeys. *Proceedings of the National Academy of Sciences*, 106(47), 20069-20074. doi:10.1073/pnas.0905314106
- 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. doi: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. doi: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. doi:10.1371/journal.pone.0089642
- 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. doi:10.1017/S1366728920000152
- Mårtensson, J., & Lövdén, M. (2011). Do Intensive Studies of a Foreign Language Improve Associative Memory Performance? *Frontiers in Psychology*, *2*(12). doi:10.3389/fpsyg.2011.00012
- 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. doi:10.1371/journal.pone.0141825
- 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. doi: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. doi:https://doi.org/10.1016/0028-3932(71)90067-4
- Panizzon, M. S., Fennema-Notestine, C., Eyler, 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. doi: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. doi: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. doi:10.1007/s00429-020-02115-5
- Qi, Z., Han, M., Garell, 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. doi: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. doi: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. doi: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. doi:https://doi.org/10.1075/eurosla.16.07rog
- Rogers, V. E., Meara, P., Barnett-Leigh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association*, *1*(1), 49-60. doi: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. doi: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. *Magn Reson Imaging*, *57*, 194-209. doi: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. doi:https://doi.org/10.1016/j.neuroimage.2019.116096
- Smith, S. M. (2002). Fast robust automated brain extraction. *Hum Brain Mapp*, *17*(3), 143-155. doi: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*, S208-219. doi: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. doi: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. doi:10.1111/j.1600-0404.1987.tb03597.x
- Team, R. C. (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. doi: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. doi: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). doi: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). doi: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. doi: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. doi:10.3758/BF03192720
- Van de Moortele, P.-F., Auerbach, E. J., Olman, C., Yacoub, E., Uğurbil, 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. doi: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. doi: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. *Eur J Radiol*, 82(5), 708-718. doi: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. doi: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. doi:https://doi.org/10.1016/j.bandl.2009.12.005
- Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of Neurophysiology*, 100(6), 3328-3342. doi:10.1152/jn.90355.2008
- Vincent, J. L., Snyder, A. Z., Fox, M. D., Shannon, B. J., Andrews, J. R., Raichle, M. E., & Buckner, R. L. (2006). Coherent spontaneous activity identifies a hippocampal-parietal memory network. *J Neurophysiol*, 96(6), 3517-3531. doi:10.1152/jn.00048.2006
- 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. doi:10.1002/mrm.26124
- Vuoksimaa, E., Panizzon, M. S., Chen, C.-H., Fiecas, M., Eyler, 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. *Cereb Cortex*, 25(8), 2127-2137. doi: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. *Med Image Anal*, 58, 101559. doi:https://doi.org/10.1016/j.media.2019.101559
- Wasserthal, J., Neher, P. F., & Maier-Hein, K. H. (2018a, 2018//). *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. doi:https://doi.org/10.1016/j.neuroimage.2018.07.070
- Wen, Z., Biedroń, A., & Skehan, P. (2016). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, 50(1), 1-31. doi: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. doi:10.1523/jneurosci.0999-11.2011
- Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., & Zatorre, R. J. (2008). Volume of Left Heschl's Gyrus and Linguistic Pitch Learning. *Cerebral Cortex*, 18(4), 828-836. doi: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. doi: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. *Cereb Cortex*, 20(3), 549-560. doi:10.1093/cercor/bhp119
- Zaretskaya, N., Fischl, B., Reuter, M., Renvall, V., & Polimeni, J. R. (2018). Advantages of cortical surface reconstruction using submillimeter 7 T MEMPRAGE. *NeuroImage*, 165, 11-26. doi:10.1016/j.neuroimage.2017.09.060

Supplementary Information

Correlation tests on cortical thickness and surface area from relevant cortical areas on each LLAMA tests. Age and estimated intracranial volume added as a covariate of no interest to all analyses.

CorticalArea	Mean Cortical Thickness \pm SD (mm)	LLAMAB correlation Pearson's r	LLAMAB correlation p	LLAMAD correlation Pearson's r	LLAMAD correlation p	LLAMAF correlation Pearson's r	LLAMAF correlation p
Left Superior Temporal Gyrus	2.30 \pm 0.19	0.263	0.055	-0.146	0.291	0.21	0.131
Left Middle Temporal Gyrus	2.27 \pm 0.19	0.159	0.25	-0.047	0.738	0.203	0.144
Left Transverse Temporal Gyrus	2.75 \pm 0.19	0.111	0.422	0.116	0.404	0.135	0.334
Left Inferior Frontal Gyrus pars opercularis	2.33 \pm 0.32	0.088	0.525	-0.287	0.036	0.126	0.368
Left Inferior Frontal Gyrus pars triangularis	2.29 \pm 0.30	0.149	0.283	-0.314	0.021	0.012	0.934
Left Supramarginal	2.73 \pm 0.13	0.099	0.478	-0.054	0.698	0.161	0.249
Right Superior Temporal Gyrus	2.66 \pm 0.29	-0.01	0.942	0.09	0.516	0.039	0.782
Right Middle Temporal Gyrus	2.29 \pm 0.36	-0.055	0.694	0.061	0.661	0.088	0.532
Right Transverse Temporal Gyrus	2.85 \pm 0.21	-0.065	0.642	-0.064	0.644	0.34	0.013
Right Inferior Frontal Gyrus pars opercularis	2.64 \pm 0.17	-0.131	0.344	-0.001	0.993	0.045	0.748
Right Inferior Frontal Gyrus pars triangularis	2.53 \pm 0.20	0.078	0.574	-0.256	0.062	0.032	0.822
Right Supramarginal	2.52 \pm 0.20	-0.072	0.605	0.07	0.614	0.104	0.458

CorticalArea	Mean Cortical Surface Area \pm SD (mm)	LLAMAB correlation Pearson's r	LLAMAB correlation p	LLAMAD correlation Pearson's r	LLAMAD correlation p	LLAMAF correlation Pearson's r	LLAMAF correlation p
Left Superior Temporal Gyrus	3910 \pm 400	0.097	0.487	0	0.998	-0.023	0.868
Left Middle Temporal Gyrus	2990 \pm 370	0.001	0.993	0.176	0.202	0.349	0.01
Left Transverse Temporal Gyrus	530 \pm 90	-0.188	0.172	-0.152	0.272	-0.062	0.659
Left Inferior Frontal Gyrus pars opercularis	1680 \pm 240	0.115	0.409	-0.049	0.727	0.099	0.481
Left Inferior Frontal Gyrus pars triangularis	1420 \pm 210	0.076	0.583	-0.115	0.407	-0.059	0.676
Left Supramarginal	4140 \pm 590	0.09	0.518	0.186	0.179	0.184	0.187
Right Superior Temporal Gyrus	3690 \pm 350	0.168	0.225	-0.034	0.809	0.033	0.813
Right Middle Temporal Gyrus	3370 \pm 360	-0.115	0.407	0.058	0.675	-0.085	0.544
Right Transverse Temporal Gyrus	420 \pm 80	0.191	0.166	0.126	0.363	-0.066	0.639
Right Inferior Frontal Gyrus pars opercularis	1390 \pm 220	0.078	0.576	0.143	0.302	0.008	0.956
Right Inferior Frontal Gyrus pars triangularis	1580 \pm 280	0.084	0.544	0.076	0.587	0.025	0.857
Right Supramarginal	3800 \pm 470	0.072	0.604	0.263	0.055	-0.202	0.147

Correlation tests from each diffusion parameter on each LLAMA tests.

Correlation coefficients, t- and p-values from correlation tests from each diffusion parameter with LLAMA B (vocabulary learning aptitude) score.

	Diffusion parameter	MD	FA	MK	AD	RD	AK	RK
Left AF	Pearson's r	0.081	0.061	0.098	0.189	0.013	0.152	0.09
	t	0.591	0.442	0.708	1.39	0.0947	1.1	0.648
	p	0.557	0.66	0.482	0.171	0.925	0.277	0.52
Right AF	Pearson's r	0.103	-0.089	0.025	0.076	0.097	-0.004	0.116
	t	0.74	-0.641	0.181	0.543	0.698	-0.029	0.834
	p	0.463	0.525	0.857	0.589	0.488	0.977	0.408
Left SLF III	Pearson's r	0.014	0.083	0.177	0.08	-0.016	0.114	0.033
	t	0.101	0.595	1.282	0.573	-0.111	0.82	0.235
	p	0.92	0.555	0.206	0.569	0.912	0.416	0.815
Right SLF III	Pearson's r	0.041	-0.142	-0.068	-0.063	0.083	-0.138	-0.055
	t	0.289	-1.028	-0.486	-0.454	0.598	-0.997	-0.393
	p	0.773	0.309	0.629	0.652	0.553	0.324	0.696

Correlation coefficients, t- and p-values from correlation tests from each diffusion parameter with LLAMA D (phonetic memory) score.

	Diffusion parameter	MD	FA	MK	AD	RD	AK	RK
Left AF	Pearson's r	0.081	-0.047	-0.247	0.087	0.094	-0.523	-0.239
	t	0.583	-0.336	-1.84	0.622	0.671	-4.38	-1.76
	p	0.562	0.738	0.0717	0.537	0.505	5.92E-05	0.085
Right AF	Pearson's r	0.196	-0.144	-0.207	0.147	0.183	-0.281	-0.202
	t	1.425	-1.038	-1.511	1.065	1.33	-2.092	-1.473
	p	0.16	0.304	0.137	0.292	0.189	0.041	0.147
Left SLF III	Pearson's r	0.083	-0.184	-0.318	0.01	0.105	-0.458	-0.318
	t	0.597	-1.34	-2.397	0.075	0.755	-3.676	-2.392
	p	0.553	0.186	0.0202	0.941	0.454	0.00057	0.0205
Right SLF III	Pearson's r	0.094	0.119	-0.063	0.246	0.001	-0.17	-0.068
	t	0.673	0.857	-0.452	1.81	0.007	-1.232	-0.485
	p	0.504	0.396	0.653	0.0762	0.995	0.223	0.629

Correlation coefficients, t- and p-values from correlation tests from each diffusion parameter with LLAMA F (grammatical inferring aptitude) score.

	Diffusion parameter	MD	FA	MK	AD	RD	AK	RK
Left AF	Pearson's r	-0.129	0.0366	-0.0428	-0.153	-0.105	-0.0966	-0.0423
	t	-0.93	0.262	-0.306	-1.11	-0.757	-0.693	-0.302
	p	0.357	0.795	0.761	0.273	0.453	0.491	0.764
Right AF	Pearson's r	-0.116	0.017	-0.079	-0.135	-0.089	-0.102	-0.169
	t	-0.833	0.123	-0.566	-0.973	-0.636	-0.729	-1.224
	p	0.409	0.903	0.574	0.335	0.527	0.469	0.227
Left SLF III	Pearson's r	0.003	0.002	0.081	-0.006	0.007	0.093	0.095
	t	0.023	0.015	0.582	-0.04	0.047	0.67	0.68
	p	0.982	0.988	0.563	0.968	0.963	0.506	0.5
Right SLF III	Pearson's r	-0.014	-0.011	-0.091	-0.032	-0.002	-0.004	-0.166
	t	-0.1	-0.079	-0.65	-0.231	-0.017	-0.026	-1.203
	p	0.921	0.937	0.519	0.818	0.987	0.979	0.235

Comparisons between thickness estimates between Novén et al. (2019) and the present study

In our previous work (Novén et al., 2019), we used different MRI sequences for obtaining anatomical MRI images for cortical thickness estimation, see table below. Perhaps of most importance is the difference in spatial resolution (1 mm³ isotropic versus 0.8 mm³ in the present study). Figure 1 shows the group mean thickness from both studies while figure 2 shows the standard deviation of the group thickness estimates from each of the studies. Cortical thickness estimates are in general slightly greater in the present study. Standard deviations within the groups seem comparable, indicating that the greater resolution does not entail less varying cortical thickness estimate.

Parameters of the anatomical MRI sequences from Novén et al. (2019) and the present study.

Study	Repetition time (T1 volume)	Echo time (T1 volume)	Flip angle (T1 volume)	Inversion time (T1 volume)	Repetition time (PD volume)	Echo time (PD volume)	Flip angle (PD volume)	Inversion time (PD volume)	Resolution of corrected volume
Novén et al. (2019)	5 ms	2 ms	6°	1200 ms	6 ms	2.5 ms	7°	1200 ms	1 mm ³ isotropic
Present	8 ms	1.97 ms	8°	1200 ms	8 ms	1.97 ms	2°	1200 ms	0.8 mm ³ isotropic

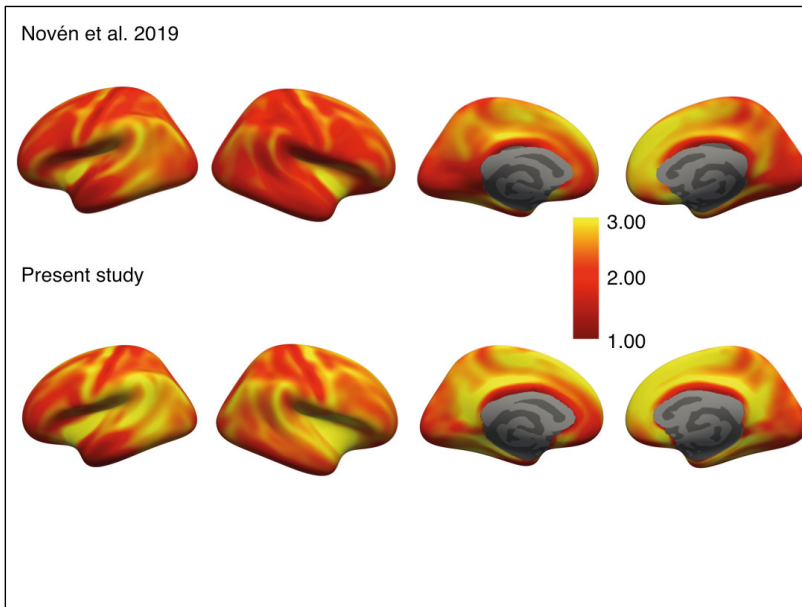


Figure 1: Group mean thickness across the cortex in the two studies.

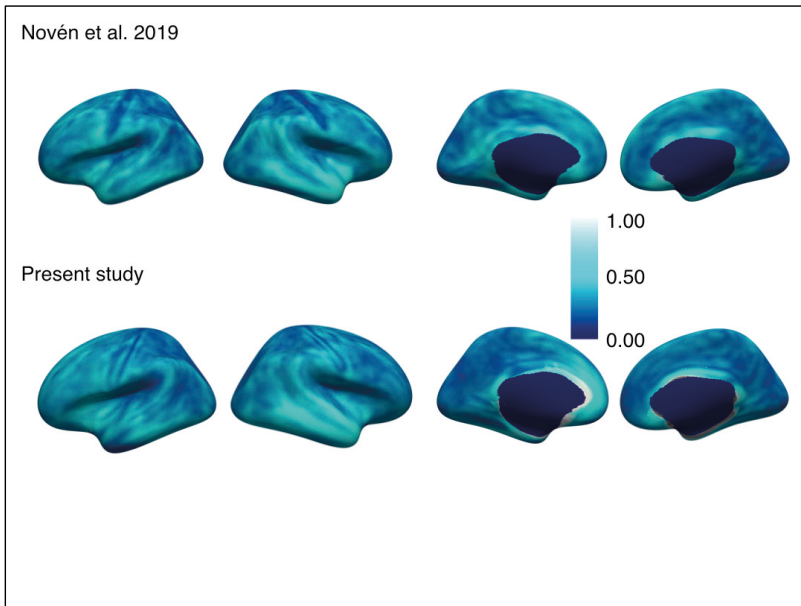


Figure 2: Standard deviations in cortical thickness estimates for the groups in the two studies.

References

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.

