Regimented language training. Changes to brain and behaviour following intensive non-native language learning

Mårtensson, Johan

2012

Link to publication

Citation for published version (APA):
Mårtensson, J. (2012). Regimented language training. Changes to brain and behaviour following intensive non-native language learning Lund University
Regimented language training

Changes to brain and behaviour following intensive non-native language learning

Johan Mårtensson

To be publicly examined in lower Palaestra, Lund University, on the 4th of December 2012 at 13:00 for the degree of Doctor of Philosophy.

Advisors: Martin Lövdén and Magnus Lindgren
Faculty opponent: Bogdan Draganski, Université de Lausanne
Examinining committee: Merle Horne, Lund University; Aki Johansson, Lund University; Sara Bengtsson, Karolinska Institutet
Copyright © Johan Mårtensson

Faculty of Social Sciences, Department of Psychology

Printed in Sweden by Media-Tryck, Lund University
Lund 2012
To Maria and Max
# Table of content

Abstract  
Svensk sammanfattning  
  Förändringar i hjärnstruktur och beteende efter intensiva språkstudier  
Acknowledgements  
List of papers  
Introduction  
Background  
  Purpose  
  The interpreter academy  
  Neural correlates of language  
    Speech recognition  
    Speech production  
    Word learning  
Measuring neural change  
  Means of analysis  
  Changes and differences  
  What are we measuring?  
I want to change my brain structure, what can I do?  
  Physical exercise and motor training  
  Non-physical training  
  What about schooling?  
  Are all results equally valid?
Objectives

Empirical studies
   Study 1
   Study 2
   Study 3
      Between groups
      Within group
   Study 4
      Between groups
      Predictive value of white matter

Discussion
   Possible implications for the aging brain
   Limitations
   Outlook

References

Table 1
Abstract

The influence of foreign-language acquisition on adult brain structure remains mostly unexplored. Whilst several cross-sectional studies have led to further understanding of how language can affect brain structure, only a few longitudinal studies have been conducted. We measured the effects of language acquisition on behaviour and brain structure at the Swedish Armed Forces Language School in two separate cohorts of army interpreters. Our hypothesis was that intensive language studies would lead to increases in associative memory performance, hippocampal volume, relevant cortical areas and connectivity between areas of the brain. In Study 1 we compared performance on 8 cognitive tasks before and after 3 months of language learning \( (n = 15) \) to a control group of university students \( (n = 19) \). Behavioural measures revealed increases in associative memory performance; remembering names and faces, a paired associates task, which is known to involve hippocampal processing. In Study 2 semi-automatic tracing of the hippocampus using FreeSurfer was compared to manual tracing and deemed equivalent to manual tracing in young adults \( (n = 44) \) but not in older adults \( (n = 47) \). Evaluation was based on data from earlier studies. Study 3 used FreeSurfer to measure grey matter volume in subcortical structures as well as cortical thickness in the next cohort of interpreters and university students \( (n = 14 \text{ and } 17, \text{ respectively}) \). Results revealed increases in hippocampal volume and cortical thickness of the left middle frontal gyrus, inferior frontal gyrus, and superior temporal gyrus. The right hippocampus and the left superior temporal gyrus, both key regions in language acquisition, increased relatively more in interpreters with higher language proficiency following training. Interpreters struggling to stay at the academy showed relatively higher increases in the middle frontal gyrus, a region that is known to be involved in processing difficult speech. In Study 4 the same group of interpreters as study 3 was measured using Diffusion Tensor Imaging. No changes in white matter microstructure were observed. A possible explanation is that interpreters had well developed tracts before starting their training at the academy. Effects in this direction could be observed as lower values of radial diffusivity in interpreters relative to controls. The combined results showed that grey matter volume in relevant language areas as well as the right hippocampus increased for interpreters compared to controls. Furthermore, increases were related to separate behavioural measures and ratings. The results indicate that grey matter plasticity is not only possible during language acquisition, but the extent and region of plastic changes varies between individuals.
Regimented language training, Changes to brain and behaviour following intensive non-native language learning

Abstract
The influence of foreign-language acquisition on adult brain structure remains mostly unexplored. Whilst several cross-sectional studies have led to further understanding of how language can affect brain structure, only a few longitudinal studies have been conducted. We measured the effects of language acquisition on behaviour and brain structure at the Swedish Armed Forces Language School in two separate cohorts of army interpreters. Our hypothesis was that intensive language studies would lead to increases in associative memory performance, hippocampal volume, relevant cortical areas and connectivity between areas of the brain. In study 1 we compared performance on 8 cognitive tasks before and after 3 months of language learning (n = 15) to a control group of university students (n = 19). Behavioural measures revealed increases in associative memory performance; remembering names and faces, a paired associates task, which is known to involve hippocampal processing. In study 2 semi-automatic tracing of the hippocampus using FreeSurfer was compared to manual tracing and deemed equivalent to manual tracing in young adults (n = 44) but not in older adults (n = 47). Evaluation was based on data from earlier studies. Study 3 used FreeSurfer to measure grey-matter volume in subcortical structures as well as cortical thickness in the next cohort of interpreters and university students (n = 14 and 17, respectively). Results revealed increases in hippocampal volume and cortical thickness of the left middle frontal gyrus, inferior frontal gyrus and, superior temporal gyrus. The right hippocampus and the left superior temporal gyrus, both key regions in language acquisition, increased relatively more in interpreters with higher language proficiency following training. Interpreters struggling to stay at the academy showed relatively higher increases in the middle frontal gyrus, a region that is known to be involved in processing difficult speech. In study 4 the same group of interpreters as study 3 was measured using Diffusion Tensor Imaging. No changes in white matter microstructure were observed. A possible explanation is that interpreters had well developed tracts before starting their training at the academy. Effects in this direction could be observed as lower values of radial diffusivity in interpreters relative to controls. The combined results showed that grey matter volume in relevant language areas as well as the right hippocampus increased for interpreters compared to controls. Furthermore, increases were related to separate behavioural measures and ratings. The results indicate that grey matter plasticity is not only possible during language acquisition, but the extent and region of plastic changes varies between individuals.

Key words
Cortical thickness, Hippocampus, Plasticity, Language acquisition, Associative memory, Transfer, Education

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 2012-10-29
Svensk sammanfattning

Förändringar i hjärnstruktur och beteende efter intensiva språkstudier

Johan Mårtensson
Institutionen för psykologi, Lunds universitet, 2012

fick kämpa relativt mer för att få stanna kvar vid utbildningen uppvisade större ökningar i mellersta frontala gyrus, ett område som är kopplat till processandet av svårbegripligt språk. Studenter med bättre skolprestation hade större ökningar i superiörta temporala gyrus samt hippocampus, två viktiga områden vid språkinlärning. I studie 4 uppvisades inga förändringar över tid i vitsubstans-struktur, något som troligtvis beror på att studenterna vid tolkskolan redan från början hade en högre integritet i trakter som angränsar relevanta språkområden. Sammantaget är slutsatsen att grå-substans kan förändras till följd av intensiv språkträning, att förändringar sker i områden som är relevanta för språkprestation och att grad och lokalisering av utveckling varierar på individuell nivå.
Acknowledgements

The work presented in this dissertation would not have been possible without the help of others.

First, I would like to thank my supervisors Martin Lövdén and Magnus Lindgren. Your faith in my ability, continuous encouragement and sense of responsibility has been a source of strength during these sometimes trying years. Martin is as skilfully adaptive as a supervisor as he is as a researcher; ensuring progression with constant but manageable challenges. Magnus has, through our many discussions, shown me the strengths and limitations inherent in any study; paving the way for understanding of what makes good research good. I could never have asked for more than the combined resources of these gentlemen.

I would also like to thank Per Johnsson and the Department of Psychology for accepting me into the PhD Programme and providing a positive work environment. Lars Kindberg and Tommy Bauer deserve separate mention for always going that extra mile. I am deeply grateful to Jean-Christophe Rohner and Magnus R. Larsson for their encouragement and support during my first years in research and continued friendship. Furthermore I would like to thank Johan Hellman for our many discussions and combined efforts, and Johan Bertlett for always believing in me.

The presented work would not have been possible without the effort of the students and staff at the Swedish Armed Forces Language School. I am especially grateful for the chance to collaborate with Stefan Kindblad and Magnus Dahnberg.

My stays in Umeå were some of the best moments during these past years due to the hospitality of Lars Nyberg and Johan Eriksson. I would also like to thank Ann-Kathrine Larsson and Hans-Olov Karlsson for their work during data collection for Studies 3 and 4. I have since been taught how to operate an MR scanner and your performance remains my golden standard.

I would like to thank Nils Bodammer for our many conversations and late night coding sessions, as well as Elisabeth Wenger for helping on numerous occasions and for being my colleague-abroad. I am grateful to Sohail Asghar for our conversations about life and research, he has wisdom far beyond his years.

I thank my friends for listening to me go on and on about brains these past few years. If you come to Berlin next year coffee is on me.
I would also like to thank Anneli Jonstrup and Aron Hejdström for making sure that we, as a family, enjoy an equal amount of board gaming and movie watching.

Any work of this scale and length will lead to moments of mental exhaustion. I thank my sister, Nina Mårtensson, for always believing in my ability, even during those moments.

I would like to thank Paula and Sven Jonstrup for keeping us in mind and helping us when needed. I am grateful to my father for his support and interest in my work. Thanks are also due to my mother for always helping me, Maria and Max during periods of intense work. We would not have managed without you.

This work is dedicated to my wife Maria Jonstrup and our son Max August Mårtensson Jonstrup. *I wish that you have all the years remaining of my life.*
List of papers

This dissertation is based on the following studies:


Paper III is reprinted with permission of Elsevier ©.
Introduction

The aim of this dissertation is to further our knowledge of the effects of non-native language learning on brain and behaviour. Understanding the functional anatomy of language has been a topic of investigation for more than 130 years (Hickok & Poeppel, 2007). Despite this, it is only recently that longitudinal studies into the effects of language learning on adults have started to appear (Mårtensson et al., 2012; Schlegel, Rudelson, & Tse, 2012; Stein et al., 2010).

If you open a ten-year-old book on neuropsychology and look for the term ‘plasticity’ chances are good that you will end up reading about children, developmental disorders, and recovery after accidents. It is less likely that you find a section on structural brain changes in adults. Being able to look at the effects of different forms of experience on the adult human brain is a fairly recent development. Measuring such effects has been made possible by neuroimaging techniques that can be used to measure brains in vivo: within the living.

Plasticity denotes the nervous system’s capacity for change. Changes are driven by experience, such as learning or thoughts, and are fuelled by discrepancy between our abilities and demands that our environment places on us. As proposed by Lövdén and colleagues (2010) plasticity will manifest as a reaction to mismatch between supply and demand. This is illustrated in a much cited study by Draganski et al. (2004). Participants were taught how to juggle. The brains of these young adults were studied before and after 3 months of practise. An increase in relevant regional grey matter volume was observed after participants had learned how to juggle. Arguably this increase was due to increased demands on these regions, demands that were larger than supplied ability. 3 months after juggling practise stopped Draganski and colleagues (2004) saw a decline. Once again one would expect mismatch between demands; a lack of demand from juggling led to a decrease in areas where supply was higher than needed. To understand when and at what level of discrepancy these changes occur is key to understanding neural development. If the brain would constantly respond to exterior demands with changes to brain structure it would likely be maladaptive (and it sometimes is, see Draganski & May (2008) for a good section on maladaptive plasticity). For changes to occur, demands for neural change should be stronger than demands for neural stability (Bäckman & Dixon, 1992). Or as Gerd Kempermann put it on page 36 in Vogel (2012) ‘A fully plastic brain is not very helpful … It learns
everything but remembers nothing’ in response to why our brain becomes less flexible as we mature.

In animals it is well known that enriched environments can lead to changes in cortical thickness (Rosenzweig, Bennett, & Diamond, 1972) or brain volume (Rosenzweig, Bennett, & Krech, 1964). Rats trained to reach for pellets showed increased dendritic branching in the motor cortex (Withers & Greenough, 1989), whilst rats trained in a motor skill task showed increased number of synapses per neuron in the cerebellar cortex (Black, Isaacs, Anderson, Alcantara, & Greenough, 1990). Van Praag, Kempremann and Gage found evidence of cell proliferation and neurogenesis in the dentate gyrus of adult mice following running practise (1999). These original findings brought hope of similar effects in humans and led to research into how experience can change brain structure in human adults (Anderson, 2011). This, in turn, led to several cross-sectional comparisons between groups with different backgrounds or levels of experience. Maguire and colleagues (2000) compared London taxi drivers to a control group and saw that they had increases in posterior hippocampal volume. The hippocampus is involved in spatial navigation and memory, with the posterior part of the hippocampus holding special importance for navigation. To become a taxi driver in London you have to be able to find and know the names of most of the streets in this very large city. This takes extensive training during 2 years in what is colloquially known as ‘The Knowledge’. With this background it is plausible that their training and experience had led to increases in hippocampal volume. Musical experience also proved to be related to brain structure; Gaser & Schlaug found that professional musicians had relatively larger brain volumes in the left precentral gyrus, right superior parietal cortex, and the left Heschl’s gyrus (Gaser & Schlaug, 2003a, 2003b) whilst Bengtsson et al. (2005) found correlations between piano practice and white matter microstructure, with childhood practice having larger effects than practising in adulthood. As any cross-sectional study, these findings run into the Chicken or the Egg causality dilemma; are skilled musicians or taxi drivers skilled musicians and taxi drivers because they have a larger capacity to handle the challenges posed by their line of work? Or has their line of work and amount of training changed their brains? To be certain that experience can change brain structure in adulthood we must measure brain structure before and after a novel experience. Several longitudinal studies have since been conducted in a number of settings, starting with juggling (Draganski, et al., 2004) and continuing with numerous forms of training as outlined in Table 1 and the section titled ‘I want to change my brain structure, what can i do?’. An example of this development is Woollett and Maguire’s (2011) follow up study on the taxi drivers in London. Results were similar to the original study (Maguire, et al., 2000) and indicate increases over time in the posterior hippocampus during navigational practice. This tells us that the larger than average posterior hippocampi of taxi drivers is an effect of experience, rather than an endowment prior to training.
Acquiring a foreign language in adulthood is a possible item on peoples to-do list. It could be to prepare for a vacation, work abroad or simply because learning a language can be challenging and fun. However, the effects of language learning on brain structure have just recently begun to unravel. This contrasts with the abundant literature on neurophysiology of native language (Demonet, Thierry, & Cardebat, 2005; Gaskell & Marslen-Wilson, 1997; Hickok & Poeppel, 2007; Price, 2010). Other types of training such as juggling (Draganski, et al., 2004) or academic studies (Draganski et al., 2006) have proven that grey matter structure in the adult human brain can change as an effect of experience. More recent studies have shown effects over time on brain structure following intervention in a range of areas: meditation (Tang et al., 2010), balancing (Taubert et al., 2010), physical exercise (Erickson et al., 2011), working memory training (Lövdén et al., 2010) as well as when acquiring (Schlegel, et al., 2012) or improving on (Stein, et al., 2010) second language proficiency.

This dissertation will investigate the effects of language training on behaviour and brain structure in a unique group of students: conscript interpreters at the Swedish Armed Forces Language School. The students at the academy consist of highly motivated and select individuals that learn a new language from scratch to near fluency in less than a year. They are roughly of the same age, in good physical shape before starting training, and study at a pace that would be almost impossible to achieve outside of their context. In many ways they present the perfect group to study; we were given the opportunity to observe how a new language is acquired over very short time and the possible effects of that acquisition on brain structure.

In the following sections I will describe the purpose of this dissertation and evaluate the current findings of changes in grey and white matter structure as well as relevant literature on the neural correlates of language.
Background

Purpose

Why spend time and resources on measuring the effects of language learning on adult human brains? By studying non-native language acquisition we can further our knowledge of how the brain is connected and built and how it responds to a fairly unique, in its intensity, learning environment at the interpreter academy. In contrast to the well mapped neurophysiology of language (Demonet, et al., 2005; Gaskell & Marslen-Wilson, 1997; Hickok & Poeppel, 2007; Price, 2010) only a few longitudinal studies have been conducted in language settings (Schlegel, Rudelson, & Tse, 2012; Stein et al., 2010). Besides measuring brain structure before and after learning, it is of importance to compare possible changes with behavioural outcomes. To quote Zatorre, Fields and Johansen-Berg (2012; page 530):

There are several reasons why it is important to establish relationships between MRI-based effects and behavior. First, in the context of learning, many different processes may be operating in parallel (for example, learning words of a new language may entail auditory discrimination, motor articulatory skills and semantic memory). If multiple brain regions show changes, interpretation is enhanced if they can be linked to separable behavioral effects. Second, individual differences in what is learned or how it is learned may not be understood without measuring behavioral outcomes. Finally, for such research to have clinical relevance, it is essential to establish correlations between behavior and structure if one wants to understand disorders that are diagnosed on the basis of behavioral disturbances.

The interpreter academy

‘Tolkskolan’, the Swedish Armed Forces Language School, is well known in Sweden for producing skilled interpreters in very short time. Most students go from no prior knowledge to fluency in an assigned language over the course of a year. Students were recruited from a subcategory of able-bodied men and women who underwent evaluation for conscription the year they finished high school. In case of women these trials are voluntary and can be performed at a later date. At the point of writing the Swedish conscription system has been scrapped and replaced by a professional army.
Both cohorts tested in this dissertation are within the old conscription system and as such represent the last of their kind. To enter the academy, students should have the highest grades in at least two foreign languages from high school. Applicants are also screened on other school achievements, study skills, emotional stability, and intelligence. The ones that pass are given one week to learn 350 non-words as a final exam before entry. Out of roughly a hundred applicants the army then chooses about 20-30 applicants per year, despite having the capacity for twice as much (see Dahlquist (2004) for reference). The reason for this is the very hard trials, which only a few individuals manage to pass. Those who do get to enter are assigned a language at the academy. Studies at the academy entail, among other things, learning between 300-500 words per week. Word acquisition is often done by means of turn cards with a foreign language word on one side and its Swedish counterpart on the opposite side, a task that is very similar to a classic paired associates task. See Paper 1 (Mårtensson & Lövdén, 2011) for description of a typical paired associates task. Taken as a whole, the interpreters study situation and selection creates a near unique learning environment. It also lends the opportunity to study how a language takes form in a controlled environment. The participants eat, live and sleep under the same circumstances. They exercise regularly and undergo the same form of training with the same expectations placed upon them.

I present data from two separate cohorts of interpreters; in Study 1 (Mårtensson & Lövdén, 2011), we had a majority of the students at the academy perform cognitive measures before and after training; in study 3 (Mårtensson, et al., 2012) and 4 we had the next cohort of interpreters undergo Magnetic Resonance measurements of grey matter volume and white matter microstructure. The first cohort studied Persian and Egyptian Arabic, the second cohort studied Dari, Egyptian Arabic and Russian. In both cohorts each student was assigned a language where they had no prior knowledge. Any effects of language acquisition are effects from learning an entirely novel language.

Neural correlates of language

What are the possible candidates for changes in brain structure following language training? In the following sections I will describe areas that are known to be involved in language processing. I will also describe the findings of two precursory longitudinal studies of language learning or experience, namely Stein et al., (2010) and Schlegel, Rudelson and Tse (2012). Focus will be on areas that are of interest for language production and comprehension as outlined by known models of language processing. Accordingly I have divided the areas into areas related to speech recognition and speech production as per Hickok and Poeppel’s dual stream model of language processing (2007) or Cathy J. Price’s categorization of fMRI studies (2010); with a
complementary account of word learning, as per Davis and Gaskell’s model of the same (2009).

According to the dual-stream model of speech processing (Hickok & Poeppel, 2007) speech recognition is handled by a ventral stream that maps sounds to meaning. Relevant areas of the brain are located in the middle and superior parts of the temporal lobe. The ventral stream maps speech input onto conceptual and semantic representations and involves several levels of processing in distributed areas in the brain. Speech production is managed by a dorsal stream that maps speech signals to articulatory networks in frontal areas of the brain necessary for speech development and production. Areas believed to be part of the dorsal stream are located in the posterior frontal lobe and the posterior dorsal areas of the temporal lobe and parietal operculum. A complementary model from Davis and Gaskell (2009) states that initial word learning leans on the hippocampus, followed by slower consolidation in cortical areas. This view is in line with earlier findings from functional MRI, where hippocampal activity has been known to differentiate good learners from bad ones (Breitenstein et al., 2005).

Speech recognition

When we hear something, be it a word or a sound, acoustic input is processed and meaningful information is extracted. The incoming information can then be translated into semantics: the meaning of messages or spellings. To manage this processing we rely on cues in the form of sounds or visual stimuli such as hand motions, facial cues or lip movements (Price, 2010).

Prelexical processing can be defined as the processing of speech up until lexical recognition of explicit words and their meaning. This includes analysis of acoustics as well as temporal aspects. Brain activation measured by fMRI show that prelexical processing occurs bilaterally, that is, in both of the brains hemispheres (Price, 2010). This is contrary to the classic view of language, where the left hemisphere dominates language processes. In (2009) Karsten Specht, Berge Osnes and Kenneth Hugdahl used a dynamic sound morphing technique to merge auditory noise into speech or music. White noise was gradually transformed into either form of stimuli. The resulting activation patterns were compared and activation was greater in the left hemisphere, especially in the superior temporal sulcus and superior temporal gyrus, for gradually appearing words compared to music. In contrast, the superior temporal sulcus and the superior temporal gyrus of the right hemisphere reacted similarly regardless if noise turned into music or words; indicating that the left hemisphere is more sensitive to speech signals compared to the right hemisphere, whilst both hemispheres are involved in prelexical phonemic processing. A phoneme is the smallest unit of sound that can be used to form the meaning of utterances.
After prelexical processing of auditory information sounds are semantically processed for meaning. A good way of identifying regions involved in speech recognition is comparing spoken non-words to spoken words. Spoken words will have a prior semantic meaning; we know what they represent. Spoken non-words consist of phonotactically legal words that initially lack meaning but sound like spoken language. An example from everyday life is the names of painkillers or other products with made-up names such as ‘Ipren’ or ‘Alvedon’. By creating non-words and pairing them with unknown objects we can create an unbiased learning situation where prior knowledge hold little influence. Another option is using material that is no longer used in today’s society: Cornelissen, Laine and colleagues (2004) had participants learn the names of old Finnish tools and their corresponding names. The names of objects was in the participants native Finnish, and hence phonotactically legal. And beautifully, the objects were actual tools, tools that no longer saw use or were described in modern language and hence would not interfere with prior knowledge. In this experiment using whole-head magnetoencephalography cortical learning was observed in the inferior parietal lobe. In three out of four subjects the cortical effect was left-sided.

So what other areas are active during semantic processing? Distributed regions around the ventral, anterior and posterior areas that border the regions that are used in prelexical speech processing (Price, 2010). Additionally, non-words seem to activate prelexical areas as well as semantic areas. A possible explanation is that processing of non-words leads to activation of regions involved in processing of known words with similar features such as shared phonemes or syllables. As such demands are higher on prelexical areas when we are presented with non-words compared to known words (Davis & Gaskell, 2009). During perception of difficult speech, ventral parts of the inferior frontal gyrus that are involved in working memory are active (Price, 2010).

Taken as a whole prelexical and semantic processing of speech is highly distributed and suggests that there are multiple pathways of language between anterior, ventral and posterior areas of the brain in both hemispheres. There also seems to be a specialization for speech processing in the left hemisphere where certain areas are more responsive to speech than other forms of acoustic input such as music (Price, 2010).

Speech production

When we want to say something we need to articulate our ideas. This requires retrieval of words that are associated with these ideas or concepts and present them verbally. This is a fairly complex task since we need to retrieve and sequence several words whilst we inhibit other, similar or interfering, words (Price, 2010). Imagine that you are proficient in both Swedish and English. This is likely to be the case for
several readers of this dissertation. Then say ‘the car is parked behind the bus’. To produce this sentence you must retrieve the relevant words and pronounce them in the correct order. Since you have representations for the word ‘car’ or ‘bus’ in both Swedish and English you also need to inhibit the representation in your other language, or other similar words such as ‘moped’ or ‘truck’. If we fail to supress our other representations we might end up with ‘the car is parked behind the buss’. Buss being the Swedish word for bus, and pronounced rather differently.

Sequencing of words on a sentence level requires getting the words in the correct order. Within a given word, sequencing is important to get the correct combination of phonemes and syllables. Syllables are units of organisation when sequencing words. For example ‘abide’ (a-bide) is a two-syllable word, whilst ‘dog’ is a one-syllable word. Pitch is also important, as is prosody. Phonemes and syllables that are to be produced must then be formed by series of movements in our laryngeal musculature (Peck et al., 2009; Price, 2010).

Conceptual processing must happen for us to be able to retrieve relevant words for speech production. Several of the regions that are involved; the inferior frontal gyrus; ventral and dorsal medial prefrontal cortex; posterior parts of the parietal lobe; the middle temporal gyrus; the fusiform gyrus and posterior parts of the cingulate gyrus overlap with areas known to be used in speech comprehension, with similar activation patterns during conceptual processing as during single-word comprehension. (Price, 2010)

Following conceptual processing words need to be retrieved. This involves activation of the left middle frontal cortex as well as the dorsal and ventral pars opercularis. (Price, 2010)

The area that traditionally has been labelled as Broca’s area is located in the inferior frontal gyrus. Broca’s aphasia is characterised by difficulties in conveying language. This area, together with premotor cortex and the anterior insula, is involved in articulation and connect to parieto-temporal areas involved in sensorimotor aspects of speech (Hickok & Poeppel, 2007). Articulation also tends to increase activation in the supplemental motor area, the superior temporal gyri, temporo-parietal cortices and the anterior insula, all bilaterally (Price, 2010). Brown and colleagues (2009) compared an overt speech task to vowel production and tongue and lip movement. The overt task consisted of reading passages of Old English heroic poem Boewulf. Participants did so with their teeth together; limiting jaw movement and unwanted activation. Results showed strong motor activation for speech in the somatotopic larynx area within the motor cortex.

Once we produce words we must monitor and correct our speech. An example would be how one might talk louder in case someone fails to hear us or if ambient noise around us increases. This requires integration of auditory, articulatory, somatosensory and motor areas (van de Ven, Esposito, & Christoffels, 2009).
It is worth noting that during speech monitoring (van de Ven, Esposito, & Christoffels, 2009), and in general, there is considerable overlap between the areas involved in production and comprehension, and although several areas are dominant in the left hemisphere many are bilateral (Hickok & Poeppel, 2007; Price, 2010).

**Word learning**

Learning new vocabulary is an important aspect of learning a new language. Vocabulary acquisition can be said to contain two parts; rapid acquisition of new words followed by slower consolidation of information. The hippocampus is believed to be involved in rapid acquisition of new words (Davis & Gaskell, 2009). Findings from functional MRI illustrate the importance of the hippocampus during word learning: In a study by Breitenstein et al. (2005) participants were shown nonsense words over 5 iterations. That is, pairs of non-words were presented, in total 5 times per combination. These word pairs were either consistent or inconsistent and in the case of consistent combinations (i.e. something that resembles learning a new language) hippocampal activation declined as an effect of presentation. This indicates that the hippocampus was involved in learning new vocabulary. The superior temporal gyrus has also exhibited activation in response to nonsense words along with left inferior frontal areas and premotor regions. The anterior part of the superior temporal gyrus along with the middle temporal gyrus show elevated activation when real words are heard, indicating their connection to words that have been consolidated (Davis & Gaskell, 2009; Hickok & Poeppel, 2007). Breitenstein and colleagues (2005) also saw activation near the supramarginal gyrus in the parietal lobe during vocabulary acquisition. Additionally, Lee et al. (2007) used white matter tractography and structural MRI scans as well as behavioural measures of vocabulary to map which areas of the brain are relevant for vocabulary knowledge. They found that vocabulary knowledge is proportionately related to grey matter density in the supramarginal gyrus. The relationship was positive, with bilateral effects showing that individuals with larger vocabulary had more grey matter. Analysis of diffusion-weighted images indicated that the posterior part of the supramarginal gyrus, where the vocabulary effect was located, was linked to anterior supramarginal gyrus as well as the anterior angular gyrus. The anterior supramarginal gyrus is involved in phonological processing whilst the anterior angular gyrus is associated with semantic processing. The posterior supramarginal gyrus was identified as the link between the two, and hence a possible link between semantic and phonological processes.

Taken as a whole the hippocampus and cortical areas in the temporal cortex are involved in acquiring new words, followed by a consolidation into cortical areas used in language processing. The interpreters at the academy learn about 300-500 words per week. This high rate of acquisition should require the use of areas in the temporal cortex as well as the hippocampus.
A wealth of studies has presented cortical and hippocampal changes following varying forms of training (May, 2011; Zatorre, Fields, & Johansen-Berg, 2012). Cortical changes to the inferior frontal gyrus have been observed in exchange students improving their German (Stein, et al., 2010), and learning Chinese can lead to extensive white matter changes in areas known to be relevant for language processing (Schlegel, et al., 2012).

Considering the numerous language areas that connect to the network displayed by Schlegel et al. (2012) and the distributed nature of language in the brain (Price, 2010) one would expect findings in other key cortical areas along with the inferior frontal gyrus (Stein, et al., 2010). Additionally, the hippocampus is proven to be plastic (May, 2011; Zatorre, et al., 2012), is involved in vocabulary learning (Breitenstein, et al., 2005) as well as rapid acquisition of new words (Davis & Gaskell, 2009). It is reasonable to expect hippocampal change during word learning.

**Measuring neural change**

Commonly used methods for measuring neuroanatomical change in vivo are Magnetic Resonance Imaging (MRI) and Diffusion Tensor Imaging (DTI). Both measures are made with an MRI scanner. MRI is a non-invasive imaging method that uses a strong electro-magnetic field to align hydrogen atoms of water molecules in tissue. Following alignment, radio frequency fields are then used to systematically change the alignment of magnetization. This magnetization results in a rotating magnetic field created by the hydrogen atoms as they return to baseline that can be detected by the MR scanner. The resulting signal can be used to construct an image of parts of the body or the brain since different tissues have different magnetic properties. T1-weighted images are commonly used for measures of grey matter morphology. T1-weighted images consist of voxels, which is a three dimensional pixel, with clear boundaries between grey and white matter (Huettel, Song, & McCarthy, 2004). DTI on the other hand, measures the direction and magnitude of water diffusion within a voxel. Since it is easier for water to diffuse along rather than across myelin sheets in the brain it is generally assumed that the direction of diffusion will reflect the direction of fibres in a voxel (Gazzaniga, 2009). This assumption works well in areas with parallel fibre organizations, but is less true for regions with complex fibre crossings and dispersion (Jones, 2010). Using diffusion values within voxels throughout the brain, it is then possible to recreate images of how white matter tracts connect different areas of the brain. DTI measures are usually described using one or more out of four separate variables; Mean Diffusivity, Fractional Anisotropy, Radial Diffusivity and Axial Diffusivity. Mean Diffusivity (MD) quantifies the free diffusion within a voxel (Beaulieu, 2002; Sen & Basser, 2005). MD is likely to be higher in large tracts or in areas with low directionality such as in the ventricles. If diffusion of
water in an open space was observed it would roughly have the shape a ball, displaying a lack of directionality. Inside the brain similar findings would indicate that there are no surrounding cell walls or tissue that disturb diffusion or hinders its movement. Fractional anisotropy (FA) measures the directionality of diffusion. FA is likely to be higher when several bundles of fibres are heading in the same direction. When water molecules are restricted in movement, as is the case between two tracts, FA will be relatively higher in that area. Picture movement of water through a garden hose; if water is restricted on all sides it will flow along those limits rather than against them. MD and FA are calculated from Radial Diffusivity (RD), which corresponds to diffusion parallel to axonal fibres, and Axial Diffusivity (AD), which corresponds to diffusion perpendicular to the same (Basser, 1995). As such RD is similar to FA and lowered RD values has been linked to loss of myelin. If lowered RD implies loss of myelin, changes perpendicular to fibre orientation (AD) might indicate axonal loss (Song et al., 2002). Taken together, measures of MD, FA, RD, and AD give us the opportunity to observe white matter microstructure measured on a macroscopic scale (i.e. voxels).

Lövdén et al. (2010) used DTI to measure the effects of cognitive training on white matter microstructure. The researchers had 20 younger adults and 12 older adults perform cognitive training over the course of 180 days. Tasks targeted working memory, episodic memory and perceptual speed. After one hundred and one 1-hour sessions the researchers found changes to the anterior part of the corpus callosum. Results showed lowered mean diffusivity and increased fractional anisotropy; indicating higher directionality in the fibres. Imagine a glass of water with water molecules diffusing freely in the middle of glass in the shape of a ball. Now image that you make the glass much taller and thinner. Diffusion is likely to go from a ball to an ellipsoid as the glass lengthens. This would increase fractional anisotropy and decrease mean diffusivity, mimicking the results of Lövdén and colleagues (2010).

Means of analysis

Image acquisition using MRI leads to large amounts of data. To draw conclusions from acquired data in scientific studies it has to be pre-processed and analysed using software. The most commonly used analysis path is Voxel Based Morphometry (VBM). This is clearly illustrated in Table 1 where 24 out of 35 studies use VBM as their method of analysis. In its simplest form, VBM calculates the chance of a given voxel being either grey or white matter (Ashburner & Friston, 2000). Originally, VBM was created to detect cortical thinning. Since then VBM has evolved into whole brain measurements, with emphasis on differences in white or grey matter structure (Ashburner & Friston, 2001). Anatomical MR images are segmented into white matter, grey matter, or cerebrospinal fluid. The brains of participants are then warped and aligned to a template image (Kanai & Rees, 2011). This is the case for all
methods described in this section, and is required for analysis. The reason for warping is that the brains of all our participants have to be in the same space and shape; otherwise we cannot compare them. A specific area of the brain of a large individual might be outside the brain of a small individual if compared directly. Once the brains of all participants are in the same shape we proceed with voxel wise analysis between groups and individuals. Although descriptions vary VBM is essentially an image analysis where the quantity measured is grey matter volume.

FreeSurfer is another software package that contains two separate pathways for analysing cortical and subcortical data. In the cortical stream FreeSurfer uses intensity and continuity information from T1-weighted images to reconstruct and measure cortical thickness across the entire surface of the brain, providing a measure of cortical thickness for any given voxel. In the subcortical stream predefined tracings of known neuroanatomical structures are used as a basis for algorithms that semi-automatically trace and return size estimates for subcortical volumes (Fischl & Dale, 2000; Fischl et al., 2002). FSL/FIRST (Patenaude, Smith, Kennedy, & Jenkinson, 2011) is similar to the subcortical stream of FreeSurfer but from the FSL software package. In this model-based segmentation tool models based on pre-segmented images from the Center for Morphometric Analysis, MGH, Boston are used as basis for computerized tracing. When looking at subcortical data the main difference between Freesurfer and FSL/FIRST compared to VBM is that these two measures use algorithms that are based on pre defined anatomical tracings from manual tracing; essentially trying to recreate the workmanship of an anatomist, compared to image analysis using VBM.

Manual tracing consists of drawing the outlines of known anatomical structures and is still considered the golden standard for measurement of subcortical structures. The method is time-consuming but highly accurate if tracers have an eye (and a steady hand) for neuroanatomy. A light controlled room is required along with a calibrated screen and high quality drawing pad. The tracer carefully inspects each slice of a participant’s brain containing the neuroanatomical structure, for example the hippocampus. The outlines of an anatomical structure are then traced (drawn) onto each slice. This is repeated for all the relevant slices of a given brain. Manual tracing is commonly done within a software package, such as Analyze (http://www.analyzedirect.com/). When all slices are combined you end up with a volume measurement of the anatomical structure in question.

Tract Based Spatial Statistics (TBSS) is another part of the FSL package that performs voxel-wise statistical analysis of FA, MD, RD or AD. This is done on a skeleton of tracts that is created by using all the voxels that are common between participants (Smith et al., 2006). If one group shows significant changes in an area of the skeleton compared to another group we can draw the conclusion that white matter microstructure in this area has changed.
Using MRI and DTI it is possible to measure local grey and white matter volume as well the connections between areas of the brain. When images are collected before and after experience it is possible to compare images as a measure of structural change.

Changes and differences

MRI can be used to measure neuroanatomical change. To study actual change, rather than differences between groups, we need to design our studies accordingly. In clinical studies it common to speak about comparisons between a group of patients and a group of healthy controls as changes in the patient group. This is based on the notion that we have well known patterns of disease. The effects of Alzheimer’s for example, are very well documented and have a rich history of conducted research (Berchtold & Cotman, 1998). Another example is that of Phenylketonuria, a metabolic disorder that causes mental retardation, seizures and other medical problems if left untreated. Children with phenylketonuria cannot metabolize phenylalanine into tyrosine. If left untreated phenylketonuria leads to severe mental retardation, if treated you have normal child development (Centerwall & Centerwall, 2000). In cases like these it is safe to say that the disorder is the cause of change in the children. It would be immoral to use a control group that received no treatment, since we know that lack of treatment will cause severe mental retardation.

In psychology it is rarely that simple. I will highlight three recent studies into the effects of language on the adult human brain for comparison. We start with a cross-sectional study by Crinion and colleagues (2009). Jenny Crinion and her collaborators at University College London and Kings College London studied regional structural differences between speakers of tonal language (Chinese/Mandarin) and speakers of non-tonal languages (Different European languages). 31 native Chinese speakers were compared to 21 Europeans and again compared to 7 Europeans that had been taught Chinese. Voxel Based Morphometry was used and two regions were identified where grey and white matter density was higher in those who spoke Chinese; the right anterior superior temporal gyrus and the left posterior region of the insula. Importantly, these effects were found both in native Chinese speakers and students of the language that were originally English speaking. On a temporal scale we have 1 measurement per individual; a snapshot of brains and behavioural measurements in the form of a verbal fluency task and a composite picture description task that were used to control for language ability in the respective languages. To quote Jenny Crinion on page 4108:

On the basis of prior studies, we suggest that the locations of these gray and white matter changes in speakers of a tonal language are consistent with a role in linking the pitch of words to their meaning.
The important part of this sentence is *on the basis of prior studies*, since Crinion and colleagues have no measurement of brain structure before acquisition of Chinese. Hence the authors have to lean on prior studies, as well as the group of English speaking students who were taught Chinese later in life, to describe the likely change in brain structure due to learning a tonal language. If you remember the early study by Maguire et al. (2000) where taxi drivers in London were compared to non-taxi drivers you find a similar situation.

In (2010) Stein et al. published a longitudinal study where native English-speaking exchange students who were learning German in Switzerland were examined once at the beginning of their stay and again roughly 5 months later. The authors found structural changes in the left inferior frontal gyrus that correlated with second language proficiency. However, unlike the elegant control-group design by Crinion et al. (2009) where we could compare brains of native Chinese speakers to non-Chinese speakers to non-native Chinese speakers, no control group was used. In this case we know that change has occurred, but we have little control over the cause of change. MRI measures such as used by both Stein et al. (2010) and Crinion et al. (2009) have inherent measurement errors (Thomas & Baker, 2012) and can change over time due to general maintenance or machine wear. In the case of Stein et al. (2010) the range between the first and second measurement was 133-224 days depending on participants. Also, participants ranged from no prior knowledge in German to 4 years of High-school German classes. In this case we have a longitudinal study, but we cannot rule out other causes of change besides language.

Schlegel, Rudelson and Tse (2012) collected DTI images from 11 English speakers who studied Chinese and 16 control participants. A median of 8 scans were collected for each participant over the course of 9 months. During this time students met 9 times per week for a total of 7.5h. They found progressive changes over time in areas commonly associated with language in the left and right hemisphere as well as the genu of the corpus callosum. Areas included the middle temporal gyrus and inferior temporal gyrus, which are involved in lexical-semantic processing; areas of the middle frontal and inferior frontal gyri that are involved in lexical-semantic production and areas of the inferior frontal gyrus and precentral gyrus that are involved in phoneme production (Glasser & Rilling, 2008). They also found changes in the planum temporale, which connects to the superior temporal gyrus (Friederici, 2009). In this impressive study of white matter connectivity we can truly talk about changes rather than differences following language learning. Participants had no background in Chinese and were taught from scratch. Seasonal effects, scanner changes or other plausible causes of change were controlled for using the 16 controls that were matched on level of education. Any change that occurred in the white matter tracts of the brains of participants is likely due to the effects of learning Chinese.

If we compare the presented studies (Crinion, et al., 2009; Schlegel, et al., 2012; Stein, et al., 2010) we see strengths and limitations based on study design. In the case
of Crinion et al. (2009) we make use of the English-speaking but Chinese-proficient group of 7 as well as references from earlier literature to make the claim that tonal based language learning effects specific brain structure. Additionally we know that native-Chinese speakers and taught-Chinese speakers share the same differences in brain structure when compared to controls. It is important to remember that this study would not have been possible in a longitudinal setting, with native-Chinese speakers having life long experience in their native language. It is not a matter of better or worse when we compare cross-sectional studies to longitudinal studies, it is a matter of which conclusions we can draw. From this author’s point of view the longitudinal study conducted on exchange students in Switzerland (Stein, et al., 2010) is much more limited in respect to what conclusions we can draw due to a lack of control of control group or prior exposure to German.

What are we measuring?

When grey and white matter increases in volume following experience, what is actually happening within the brain? In the case of T₁-weighted images we know that the number of grey matter voxels have increased. In Diffusion weighted images from DTI we know that changes is related to white matter microstructure, following for example balancing practice (Taubert, et al., 2010). But what are we actually measuring? What is the underlying cause for these changes?

The short answer is, we do not really know for certain. There are however, several possible explanations for underlying change in grey and white matter.

Observed changes to structural plasticity can be broadly divided into three categories: changes to grey matter (neurons), changes to white matter (helper cells that provide for and connect the grey matter) and extra-neuronal change (i.e. blood vessels, changes in cell size in white matter). Grey matter changes may include neurogenesis (birth of new neurons), gliogenesis (increase in helper cells), synaptogenesis (creation of new synapses), and change in size.

Neurogenesis is the birth of neurons. There is evidence that neurogenesis occurs during learning. In rats we know that spatial learning accelerates maturation of dendritic trees in new neurons and helps them integrate into functional networks in the hippocampus (Tronel et al., 2010). Loss of adult-born neurons decreases memory performance in mice (Deng, Saxe, Gallina, & Gage, 2009). Neurogenesis seems to hold importance for memory formation. However, the amount of new neurons created in the hippocampus is relatively low compared to the total size of the same. It is likely a minor factor in the changes we can observe using magnetic resonance imaging, which is on the scale of cubic millimetres (Aimone, Wiles, & Gage, 2009; Zatorre, et al., 2012).
Gliogenesis, the birth of glia cells, occurs during learning. Glial plasticity is believed to important for aiding in the organisation and interaction between neurons and glia cells (Dong & Greenough, 2004) and could serve as a possible explanation of the effects we see when structural increases are observed using MRI (Zatorre, et al., 2012).

Synaptogenesis has been observed following motor learning, along with changes in dendritic spine morphology. Remodelling of dendritic spines has been shown to correlate with learning and behavioural improvement in mice with a small number of new spines being preserved following consolidation of new knowledge. The findings by Draganski et al. (2004), which saw cortical increases in new jugglers, increases that declined after 3 months of rest, correspond well to observed glial changes in animals. (Zatorre, et al., 2012)

One explanation for changes in grey matter is that of vasculature. Physical activity is believed to alter vasculature in the brain, with increases in vascular volume in the cerebral cortex and cognitive ability in monkeys following physical exercise (Zatorre, et al., 2012). In an innovative study, Pereira et al. (2007) found that exercise in mice primarily affected cerebral blood flow in dentate gyrus of the hippocampus, which is a known site of adult neurogenesis. The results were then carried over to humans by means of cerebral blood flow measures in the hippocampal formation in exercising adults, where similar effects on the dentate gyrus were found. The effect in mice in turn, correlated with neurogenesis in the dentate gyrus. In humans the increase of blood flow correlated with aerobic fitness and cognition.

In white matter, changes are likely attributed to number of axons and their diameter, trajectories and branching as well as package-density of fibres and increases or decreases in myelination. Extra-neuronal changes consist of angiogenesis (new blood vessels) and changes in glial cell size as well as changes to the number of glia (Zatorre, et al., 2012).

Oligodendrocyte progenitor cells retain the ability to divide in the adult brain and remain in substantial numbers in adulthood. These cells participate in repair after damage to myelin, but could participate in remodelling following learning. Parts of the brain such as the frontal lobes of the cerebral cortex myelinate late in life (Zatorre, et al., 2012) and one third of oligodendrocyte progenitor cells originate after adolescence in mice (Psachoulia, Jamen, Young, & Richardson, 2009). Can myelin change as an effect of activity? Rueegg, Kakebeeke, Gabriel and Bennefeld (2003) studied conduction velocity of action potentials in cosmonauts. 6 cosmonauts and 6 bed rest subjects underwent electromyographic measurements of the soleus nerve. Results showed that conduction velocity was lowered during bed rest, with recovery to normal levels within 10 days of ending bed rest. A similar study of hind limb unloading in rats showed myelin reduction and reduced axon diameter in the soleus nerve (Canu, Carnaud, Picquet, & Goutebroze, 2009). These results indicate
morphological changes in myelin that is activity dependent, even if the number of fibres did not change noticeably.

In the hippocampus, mossy fibres have been known to sprout following long-term potentiation (Adams, Lee, Fahnestock, & Racine, 1997) or physical exercise (Toscano-Silva et al., 2010). Pruning of sprouts is guided by activity-dependent competition; not unlike apoptosis in the developing brain. Sprouts that fail to connect to functional networks are removed. Re-routing of connections has also been observed in skill learning or repair from damage in monkeys (Johansen-Berg, 2007). How would changes like these present themselves in images acquired using MRI? In T1-weighted images you would likely spot changes, given that they are large enough, as changes to the size of a given structure. If vascular changes enlarge the hippocampus it will appear larger in images. We have no real way of knowing if the change is due to vascular changes, neurogenesis or gliogenesis however. In the case of white matter measurements using DTI it is more complicated. DTI is sensitive to different changes in microstructure. But exactly how it will respond is less clear. Many factors can influence FA or MD values, with axonal membranes as the primary determinant of diffusion anisotropy (Beaulieu, 2009). DTI can pick up on changes in myelin such as axonal membrane circumference and myelin area (Concha, Livy, Beaulieu, Wheatley, & Gross, 2010) or packaging density (Takahashi et al., 2002). In Concha et al. (2010) extracted tissue was compared to DTI measures for validation and clarification. In Takashi et al. (2002) the very large axons in sea lamprey, which are well known, were measured. In this latter study diffusion was found to be isotropic inside large axons but anisotropic in bundles of axons. These findings support the notion that the primary sources of diffusion anisotropy are cell membranes.

I want to change my brain structure, what can I do?

There are several training methods that can induce behaviourally specific neural change. In the following sections I will describe successful intervention methods that have lead to changes in grey and/or white matter. For an overview of longitudinal studies see Table 1.

Physical exercise and motor training

Cardiovascular fitness is positively associated with performance on intelligence measures (Åberg et al., 2009). Physical activity also causes increased neurogenesis in the dentate gyrus of the hippocampus, which might mediate effects of exercise on cognitive ability (Van Praag, 2008). In mice voluntary exercise leads to enhanced
survival of new-born neurons in the hippocampus (Van Praag, Christie, Sejnowski, & Gage, 1999). This effect is specific to the hippocampus with no known increases in neurogenesis in the olfactory bulb (J. Brown et al., 2003), which is the other area alongside the subventricular zone that produces neurons in mammals. The olfactory bulb responds to odour-enriched environments instead (Rochefort, Gheusi, Vincent, & Lledo, 2002). In humans aerobic fitness training has led to increases in hippocampal volume (Erickson, et al., 2011) and parts of the human cortex: the anterior cingulate cortex, superior motor areas, and the left superior temporal lobe (Colcombe et al., 2006). That physical fitness can lead to increases in hippocampal volume is important for studies of populations such as the military interpreters. One has to make sure that an increase in hippocampal volume is not confounded by physical exercise and then attributed to language learning (in our case).

The first longitudinal study on structural changes in humans was published in Nature in 2004 (Draganski, et al.). In this study young adults were divided into groups of trainees and controls. The trainees were then taught how to juggle. All participants were scanned using MRI before training and after showing proficiency. Proficiency was defined as being able to juggle for 60 seconds. Additionally, trainees were scanned again 3 months after training stopped. Brains of jugglers showed grey matter increases in mid-temporal areas as well as the left posterior intraparietal sulcus. This seminal study paved the way for numerous studies into the effects of juggling (as well as other forms of training) on the human brain. Boyke et al. (2008) replicated the findings in the mid-temporal area in older adults and found additional increases in the left hippocampus and the nucleus accumbens, whilst Driermeyer et al. (2008) found effects in the mid temporal & frontal areas as well as the cingulate cortex after as little as 7 days of training in young adults. Furthermore, Scholz et al. (2009) used DTI, as well as T1-weighted images, on jugglers in the first study showing longitudinal changes in white matter in humans. In this case grey matter increases were reported in the medial occipital and parietal lobes with related white matter increases in fractional anisotropy (FA; high FA is usually interpreted as high directionality in fibre bundles) in the parieto-occipital sulcus.

Other forms of motor training have also reported effects, such as dynamic balancing on a suspended plate for 45 minutes once a week for 6 weeks. Participants that underwent this training regimen ended up with increases in motor areas as well as in the orbitofrontal cortex and underlying white matter structure (Taubert, et al., 2010; Taubert, Lohmann, Margulies, Villringer, & Ragert, 2011). Practising golf has also been shown to induce increases in cortical volume in sensorimotor areas (Bezzola, Mérollat, Gaser, & Jäncke, 2011). Different variations of hand based training following immobilization in clinical populations (Granert et al., 2011; Langer, Hänggi, Müller, Simmen, & Jäncke, 2012), 4 days of left hand training (Hamzei, Glauche, Schwarzwald, & May, 2011) and 5 days of pinching (Gryga et al., 2012)
have led to changes in various motor areas as well as changes to FA values in the right motor cortex and left corticospinal Tract (Langer, et al., 2012).

Non-physical training

A fairly recent study by Tang et al. (2010) showed increased FA in the anterior cingulate cortex following one month of integrative body-mind training, a form of meditation. No grey matter changes were found, in contrast to results from a study into the effects of Mindfulness-Based Stress Reduction, another form of meditation (Hölzel et al., 2011). In this study, grey matter changes were found in the left hippocampus, cingulate cortex, the temporoparietal junction, and the cerebellum.

The most famous example of navigational performance is a cross-sectional study on taxi drivers in London from 2000 (Maguire et al.). Taxi drivers in London are well known for their extensive training where they must be able to find thousands of places by heart. When compared to controls, taxi drivers exhibited larger posterior hippocampi, with individual volumes correlating with time spent driving a taxi. In a recent longitudinal follow up study (Woollett & Maguire, 2011), hippocampal size was shown to change over time in taxi drivers.

Navigation in a virtual zoo has also led to changes of hippocampal volume as well as cortical thickness. In a recent study, groups of young and old adults were trained in navigating a virtual zoo for four months. Change in hippocampal volume was found for trainees relative to controls (Lövdén et al., 2012) as well as in the left precuneus and the paracentral lobule in young participants only (Wenger et al., 2011).

Tetris has been known to increase cortical thickness in temporal areas (Haier, Karama, Leyba, & Jung, 2009) whilst computerized car games have led to microstructural changes in the hippocampus and parahippocampal area measured with DTI (Sagi et al., 2012). In the latter study increases were shown after as little as 2 hours of video gaming. These very quick changes indicate that brain structure is flexible and stand in stark contrast to the old view of an unchanging brain in adulthood.

Memory based practice has also been effective in changing white matter microstructure and grey matter volume. Engvig and colleagues found increases in grey matter in the right insula and the lateral orbitofrontal and fusiform cortices (Engvig et al., 2010); with reduced decline of FA values in frontal areas (Engvig et al., 2011) following verbal memory training over the course of 8 weeks. Lövdén and colleagues (2010) used cognitive training as an intervention for younger and older adults over the course of 180 days. The training battery consisted of working and episodic memory as well as perceptual speed tasks. They found increases in FA and decreases in mean diffusivity (MD; which is interpreted as the amount of diffusivity in a voxel,
What about schooling?

Learning new material includes the use of several different areas in the brain. The medial temporal lobe forms and consolidates new memories; the prefrontal cortex encodes and retrieves information; the temporal cortex, besides its specific involvement in language learning as mentioned earlier, stores knowledge (Gazzaniga, Ivry, & Mangun, 2002). The medial temporal lobe consists of several regions, such as the hippocampus and the adjacent perirhinal, entorhinal and parahippocampal cortices. These areas are central for mnemonic processes and work along with the cortex to establish and maintain long term memory in animals and humans. Damage to the temporal lobe generally results in forgetfulness, and memory is impaired regardless of which sensory modality the memory was represented in. Damage to the medial temporal lobe leaves immediate memory intact; implying that the structures are mainly involved in long term memory and storage. Amongst the presented structures the hippocampus acts like a hub that combines information from multiple sources. Tasks that require integration, such as combining information from multiple sources to form memory of events are dependant on the hippocampus. The same is true for associative memory tasks such as remembering a name and a face (Squire, Stark, & Clark, 2004). It has been suggested that rapid learning of sparse representations is done by the hippocampus and the remaining medial temporal system, whilst slow consolidation of distributed representations is consolidated by the cortex (Davis & Gaskell, 2009).

Wouldn’t one expect academic studies to have effect on brain structure in the medial temporal lobe or relevant cortices? Well they can; Draganski and colleagues (2006) studied German medical students that were preparing for a preliminary exam known as “Physikum”. This large exam is usually performed following 2 years of studies and entails 3 months of daily study sessions for the students. Medical students were measured before the test, after the test, and 3 months later. Medical students showed increased volume in the right posterior parietal cortex as well as both the hippocampi (with a larger effect in the right hippocampus). Two recent studies have explored the effects of acquiring or improving on German as a 2nd language (Stein, et al., 2010) and non-native language acquisition in young adults (Schlegel, et al., 2012), their findings will be discussed under paper 3 and 4 respectively and were discussed in section ‘Changes and differences’. Aside from these findings there is just one more study (Ceccarelli et al., 2009) that has used formal schooling as intervention method.
Are all results equally valid?

As you have seen there are numerous studies reporting changes in grey and white matter volume following experience. The path from image acquisition to final results is long, and differs from study to study. First, MR images have to be optimized for the study in question. One has to decide on areas of interest in and if measurements should focus on elements of diffusion or grey matter structure and then set MR scanners accordingly. Choice of software will also affect our output. VBM will spot changes in white and grey matter volume across the brain whilst FreeSurfer will trace areas of the brain using a priori knowledge of human neuroanatomy. Once pre-processing is done, using a choice of software, material must be analysed statistically. This entails several choices regarding statistical thresholds and is limited by methodological decisions made when planning your study. An overview of different techniques used by the longitudinal studies presented in this dissertation and more can be seen in Table 1. This table is inspired by and partially adapted from a recent review by Thomas & Baker (2012).

Kriegskorte, Simmons, Bellgowan & Baker (2009) argue that ‘double dipping’ in neuroscience can lead to distorted results. If the same dataset is used to select areas of interest and statistical testing it can bias your data. In many cases there is a control group but it is only partially used. Instead of measuring interaction between group and time with the two groups treated on an equal level it is common to first measure effects within the experimental group and apply areas of significant change as regions of interests on controls. Data is then exported from both groups, and if the resulting interaction (in pre-selected areas) between groups is significant results are presented as selective increases for the training group. This introduces selection bias and can lead to spurious effects (Kriegeskorte, et al., 2009). Out of the 34 papers presented in Table 1 only 14 studies compare groups over time on an equal basis. That number includes Study 3, which you will soon read.

One important consideration in any longitudinal study is reliability. In MRI research this is especially important since changes to software or hardware in the scanner can induce changes that appear as changes in participants brains. Other non-scanner related effects such as natural biological aging, seasonal effects or other events that affect your participants must also be taken into consideration. Because of this it is important to control for such effects by means of a control group. Preferably the control group should be scanned roughly around the same time as the group in training. Regardless of this there are several published studies that have no control group. This limits the conclusions that can be drawn.
Objectives

Will intensive language studies lead to changes in brain structure and associated behaviour? This seems likely based on the wealth of studies exhibiting changes following experience. The inferior frontal gyrus increased along with second language proficiency in exchange students in Switzerland who were improving their German (Stein, et al., 2010). Although the study in question is limited by its lack of control group and other factors that are outlined in Study 3, the results seem plausible. The inferior frontal gyrus is an important part of the articulatory network, which is important when learning to speak a foreign language (Hickok & Poeppel, 2007). In the case of Stein et al. (2010) students had varying experience in German. What areas will change during acquisition of an entirely new language? Some hints can be found in the findings of Schlegel, Rudelson and Tse (2012). Students learning Chinese from scratch showed widespread changes in white matter connecting areas commonly associated with language. These areas included the middle temporal gyrus and the inferior temporal gyrus, believed to be involved in lexical semantic processing. Areas linked to phoneme production, such as the inferior frontal gyrus and precentral gyrus also changed, as did areas involved in lexical-semantic processing such as the middle frontal and inferior frontal gyrus. Additionally changes were found in the planum temporale, which connects to the superior temporal gyrus, an area that integrates the dorsal and ventral streams of Hickok and Poeppels (2007) model of the cortical organization of speech processing. Schlegel, Rudelson and Tse (2012) provide an insight into the effects of language learning on white matter; adjacent grey matter remains an open book.

During speech production we make use of auditory, articulatory, somatosensory and motor areas, which overlap with areas used in speech perception. During perception ventral, anterior and posterior areas are used in widely distributed networks of language that probably involves multiple pathways of language (Price, 2010). During word learning, cortical areas of the temporal lobe and the hippocampus work together for initial acquisition and later consolidation (Davis & Gaskell, 2009). The hippocampus in particular has shown importance in differentiating good from bad learners of new vocabulary (Breitenstein, et al., 2005). The hippocampus has also been known to change during intensive study periods (Draganski, et al., 2006) and is subject to structural change during learning and exercise (Zatorre, et al., 2012). With this in consideration it seems highly likely that hippocampal volume will change.
following intensive language learning and perhaps relate to learning outcomes of the same.

This brings us to the Zatorre et al. (2012) quote in the beginning of the dissertation. Many different areas can be operating in parallel during language learning. If multiple brain regions can be expected to change our understanding and interpretation of these changes is enhanced if they can be linked to separable behavioural effects. To increase our understanding of adult language acquisition we must evaluate effects on behaviour along with changes to both grey and white matter using longitudinal measurements. By doing so we can hope to provide an insight into how the brain is affected by language learning in grey matter, white matter and behaviour on an individual level.

In Study 1 (Mårtensson & Lövdén, 2011) we used behavioural measures based on those used by Lövdén et al. (2010) but adapted for this highly select group of language students. We then evaluated behavioural findings, which included increases in associative memory performance; a decision was made that the likely neural substrate would be the hippocampus, and that further measures should focus on hippocampal change. We then choose a suitable method based on prior experiences in semi-automatic processing of grey matter (Study 2), and applied these methods on longitudinal data from the next cohort of interpreters at the academy (Study 3; Mårtensson et al. (2012)). In Study 4 we used DTI measurements collected from the second cohort to investigate whether language training could induce changes in connectivity and white matter microstructure in experienced language students.
Empirical studies

Study 1

Cognitive training based on working memory (Dahlin, Neely, Larsson, Bäckman, & Nyberg, 2008; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Olesen, Westerberg, & Klingberg, 2003) or executive control (Karbach & Kray, 2009) have reported success in producing transfer effects; that is, increases in one ability from training of another. Notably Jaeggi and colleagues (2008) found transfer from a demanding working memory task to fluid intelligence. The working memory task consisted of a double n-back task with both visual and auditory stimuli. Participants were simultaneously presented with auditory stimuli in the form of letters, such as ‘C, P, C, T’, and visual stimuli in the form of a box that changed location spatially. The n in n-back constitutes the distance of presentation. Consequently, in a 1-back task participants are asked to determine whether the preceding item is the same as the item presented just before by responding ‘yes’ or ‘no’. This is commonly done using response buttons, but can be performed verbally. In the case of auditory stimuli this would be ‘C’ followed by ‘C’, which would amount to ‘yes’. In a spatial version ‘yes’ would be the correct response if a box appears in the same location twice. A 2-back task is more difficult since the goal is to determine whether the presented letter ‘C’ is the same as the letter presented two trials ago. Using the example above ‘C, P, C’ the last ‘C’ would warrant a positive ‘yes’ response. 3-back, 4-back and so forth make the task more and more difficult; we need to keep more and more information in our working memory at once. In the case of Jaeggi et al. (2008) an adaptive n-back task was used, where n was incremented or decremented depending on how well the participant was doing. 4 different groups of trainees (34 people in total) were trained for different amounts of time; 8 days, 12 days, 17 days or 19 days. Each training session consisted of roughly 25 minutes of adaptive n-back training. All trainees along with 4 control groups, one for each training group (total of 35 people) were measured on standardized fluid intelligence tests of visual analogy problems; such as Raven’s Advanced Progressive Matrices (Raven, 2000) where participants are presented with a test pattern and are asked to identify a missing element. Jaeggi and colleagues (2008) found significant increases in measured fluid intelligence compared to controls and a linear relationship between increased fluid intelligence and time trained when the different training groups were compared. Working memory transfer effects have come
under scrutiny, (Shipstead, Redick, & Engle, 2012) as findings are usually based on single measures of abilities of interest and consistently measuring tasks that are close to the task trained. Finding an increase in a single cognitive task (such as fluid intelligence) does not equate a global increase in intelligence, especially if training transfers to a task that is very near the form of training. To put it in another way, the results of Jaeggi et al (2008) could be interpreted as n-back training having an effect on the performance on Raven’s Advanced Progressive Matrices (Raven, 2000) rather than an effect on fluid intelligence.

Despite these inadequacies, associative memory training has still been largely unsuccessful in producing transfer effects (Lövdén, Bäckman, et al., 2010). Considering that the key region for associative memory is the hippocampus, which is known to be adaptive under strain (Draganski, et al., 2006), this is surprising. Rate of transfer has been known to be a function of time spent in training (Jaeggi, et al., 2008; Schmiedek, Lövdén, & Lindenberger, 2010) and one would expect that with enough training, demands on associative memory in one modality (language) should transfer to cognitive performance in other, nearby tasks.

Studies at the interpreter academy have a strong focus on acquiring large amounts of new vocabulary; a task that we hypothesised should transfer to other tests of associative memory performance and possibly other cognitive measures. A total of eight measures were used. Three measures targeted associative memory performance, and could be considered near transfer from the interpreters vocabulary practise. The associative memory tasks consisted of a word-non-word cued recall task where participants were presented with Swedish words and Swedish non-words in pairs, and a word-word task where the combinations consisted of two Swedish words presented together. In both cases the trial task required the participants to fill in a missing word in a presented pair. These tasks, along with the face-name tasks described below, were meant to be similar to vocabulary learning at the academy. Working memory performance was measured in the form of a numerical n-back task (3 and 4 back) and a related alpha span task. We wanted to explore if switching between languages, a working memory intensive task (Darò & Fabbro, 1994), would lead improved working memory performance. Additionally we included two tasks of strategic episodic memory and one reasoning task as control tasks. We expected similar performance on these tasks from interpreters and controls.

By using a battery of 8 cognitive tasks and a group that associates 300-500 foreign words to their Swedish counterparts per week, we hoped to see whether learning vocabulary could lead to increases in associative memory performance.

Our findings were generally positive. We found a significant increase in a face – name cued recall task (see Figure 1) that depends heavily on associative memory. In this task participants were presented with pictures of faces along with their respective names. Later on they were presented with the same faces again but had to write the associated
names by themselves. Put into an everyday example this task is similar to shaking hands and presenting yourself to 44 (the number of trials) different people and trying to remember their names. It is quite hard. Interpreters saw a considerable 28% increase in the amount of names learned after training at the academy. The face-name task is the measure (with acceptable psychometric properties, see Paper 1) that most closely resembles vocabulary training at the academy. As such our associative memory findings are in line with findings from working memory training. Results are also in line with the critique by Shipstead, Redick and Engle (2012); our results are limited to near-transfer and to one out of eight tasks.

![Face - Name](image)

**Figure 1.** Number of correctly recalled names for interpreters and controls before and after 3 months of training. The group by time interaction is significant: $F(1,32) = 5.16$, $p = 0.03$, $\eta^2 = 0.14$. Error bars represent SEM.

Our conclusion was that associative memory performance on this specific task had changed following training at the academy. Considering that one of the neural substrates of associating names and faces is the hippocampus (Sperling et al., 2001); and that hippocampal activation during word learning can differentiate between good and poor learners (Breitenstein, et al., 2005) and earlier findings of hippocampal plasticity in medical students (Draganski, et al., 2006) we decided to focus on finding a solid measure of hippocampal volume for use in our follow up study.
Study 2

We evaluated different measurement methods that were available for measuring the hippocampus. Although Voxel Based Morphometry (VBM) has arguably been the most commonly used method in recent years the golden standard is undoubtedly manual tracing. When performing manual tracing an individual undertakes the mind-numbing task of tracing every slice of an MRI volume by hand. An image of the brain acquired using MRI, although 3-dimensional, is made up of thin slices, each slice consisting of a separate image of the brain at a given place. If your voxels are 1x1x1mm for instance, each slice will be 1 mm thick. Additionally, the room needs to be properly dark for image contrast and solid anatomic knowledge is needed. Manual tracing is the research equivalent to spending months in a submarine. To complicate matters another tracer must process all of the images once more to provide a comparison; otherwise human bias cannot be ruled out. We wanted a method that was as close to manual tracing as possible but without the required effort.

Recent software developments have made it possible to avoid manual tracing and instead rely on computerized algorithms to do the job. Three methods dominate: VBM, FSL/FIRST, and FreeSurfer. VBM is the most commonly used method and as mentioned in the section on ‘Means of analysis’ its simplest form consists of a brain wide comparison of individual voxels (Ashburner & Friston, 2000). It has been criticized for inaccuracy and for being susceptible for changes in the processing pipeline (Bookstein, 2001; Thomas et al., 2009), but remains very popular. VBM has seen continuous development and has been improved upon in recent years. Both FSL/FIRST and FreeSurfer use predefined areas (from manual tracing, originally) to perform automatic segmentation of different areas in the brain. Both methods have been compared directly to each other and to manual tracing (Morey et al., 2009); FSL/FIRST has certain benefits when tracing the Amygdala, FreeSurfer shows better accuracy and correspondence to manual tracing of the hippocampus.

In Paper 2 we compared manual tracing to FreeSurfer on data collected at the Max Planck for Human Development in Berlin. Data from this study was pre-existent and the sample and design of the study has been reported elsewhere (Lövdén, et al., 2012; Wenger, et al., 2011). We found that whilst FreeSurfer was less reliable than manual tracing for older adults, a finding that likely relates to the predefined image used by the software providers, it was generally quite good at tracing the hippocampal volumes of young adults. FreeSurfer continuously overestimated volume size as compared to manual tracing. Since this happens equally for all young participants and our study was going to measure differences over time for interpreters compared to a control group of the same age this was a non-issue. We concluded that for young adults FreeSurfer could be regarded as a reliable and valid method for measuring hippocampal volume.
Study 3

Can intense language training change brain structure in adulthood? By using MRI along with several behavioural measures we hoped to find out. A group of highly motivated conscript interpreters that had managed to pass the very high requirements for entering the academy were compared to young university students at Umeå University. Most of the students studied medicine or cognitive science and were at the beginning of their studies at the university. The choice of control group was pivotal since they underwent training that was nearly comparable to the training group of the only known study of hippocampal change following abstract learning (Draganski, et al., 2006). Considering selection bias and our quasi experimental design (lacking randomization into control or experimental group) we decided that it was more important to have a comparable control group that was as close as possible to interpreters, even if this meant that we ran the risk of them being too close to the interpreters situation.

For this reason we controlled for age, years of education, intelligence and emotional stability as described in Paper 3. Emotional stability was mainly added to ensure that interpreters, who had just undergone basic military training, were unaffected by the stresses that basic training put on young army recruits. An example from post traumatic stress research shows that intense stress is connected to lowered hippocampal volume (Bremner et al., 1995). We wanted to make sure that no individual became very stressed by basic training. Another possible factor that could confound hippocampal increases is exercise; see section ‘Physical exercise and motor training’, Van Praag (2008) in particular. Basic training in the army is physically demanding, much more so than the physical exercise recruits get during their stay at the academy. Hence the interpreters should have a lower level of physical fitness at post-test compared to pre-test; which we regretfully have no measure of at time of writing. This would go against our hypothesis, and therefore was considered fine. Otherwise one could assume that since physical training can lead to increases in hippocampal volume (Van Praag, 2008), we might falsely attribute increases in hippocampal size to language learning, when it was actually due to physical exercise.

We used MRI to measure T₁-weighted images to measure grey matter before and after 3 months of studies at the interpreter academy and Umeå University (in case of controls). Within the group of interpreters we also acquired ratings of language proficiency (Proficiency) and the amount of effort needed to stay at the academy (Struggle); we wanted to compare structural changes within the group to behavioural outcomes. As such I have divided our results into two categories, between groups and within group.
Between groups

**Figure 2.** Compared to controls interpreters showed increases in 3 cortical areas as well as the right hippocampus. 1A shows cortical increases on an brain inflated for presentation reasons since both gyri and sulci become visible; there were no decreases for either group at $K$ (cluster size, the amount of voxels within a area) = 100, $p < 0.001$. 1B shows volume change in mm$^3$ for the right and left hippocampus for both groups. The interaction between group and time for hippocampal volumes was significant: $F(1, 29) = 5.83$, $p = .002$. Hemisphere did not significantly modulate the observed hippocampus increase; $F(1, 29) = 2.92$, $p = .098$. 1C shows change of cortical thickness in the presented areas in mm$^2$. Decreases for the control group were not significant.
When compared to students as Umeå University interpreters at the academy exhibited increases in three cortical areas as well as the hippocampus. The cortical increases were in the left hemisphere whilst the hippocampal increase was right dominant. See Figure 2 for illustration.

All the cortical areas showing increases are language-related and contained within the fronto-temporal areas of the left hemisphere. The inferior frontal gyrus (IFG) and superior temporal gyrus (STG) are well-known language areas, historically known as Broca’s and Wernicke’s area respectively. They are involved in various language related tasks (Davis & Gaskell, 2009; Demonet, et al., 2005; Hickok & Poeppel, 2007; Price, 2010), with the inferior frontal gyrus being important in the articulatory network whilst the superior temporal gyrus is involved in acoustic phonetic processing as well as integrating the dorsal and ventral streams of language (Hickok & Poeppel, 2007). The middle frontal gyrus (MFG) is part of the articulatory network (Hickok & Poeppel, 2007) and is believed to be involved in processing of difficult speech (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007). The right hippocampus has been reported to be selectively involved in associative learning (Henke, Weber, Kneifel, Wieser, & Buck, 1999) and the role of the hippocampus in rapid learning of new words is predicted by Davis and Gaskell (2009).

Summary: we found selective increases for interpreters in areas that are related to language acquisition with no declines in either group.

Within group

What about the group of interpreters, was there individual variation in structural change when performance and effort were under consideration? In an attempt to evaluate if changes in grey matter structures were driven by either language proficiency (or knowledge, to put it differently) or the effort needed to perform according to the high standards of the academy (which in this case is related to reaching the minimum grades needed to be allowed to stay and study at the academy) we exported values from all the above areas for each interpreter and correlated them to values of Proficiency and Struggle. You can see the results in Figure 3 below.
There was individual variance in which areas increased within the interpreter group. 3A and B: interpreters who had a harder time achieving the goals of the academy had relatively higher increases in the middle frontal gyrus, which has been tied to processing of difficult speech. 3A, C and D: interpreters with higher language performance at the end of 3 months of studies had relatively larger increases in the superior temporal gyrus and the right hippocampus. Note the near disassociation between Struggle and Proficiency in the right hippocampus (3A), the negative trend towards significance between hippocampal change and Struggle is at \( p = -.07 \). Stars (*) indicate significant correlations (\( p < .05 \)).

These results are partially inconsistent with a recent study by Stein et al. (2010), which is the only other longitudinal language study that we are aware of that has seen changes in gray-matter structure in relation to performance. Stein et al. report changes in gray-matter density in the left inferior frontal gyrus in exchange students learning German in Switzerland. In our study we detect changes to the left IFG when compared to controls, but we see no relation between our measure of Proficiency and the left IFG. Instead we find the left superior temporal area to be a stronger determinant of language proficiency. However, the studies are dissimilar on several
important accounts and direct comparison cannot be made between the findings. In
the study reported by Stein and colleagues (2010) prior knowledge of German varied
between participants as compared to no prior knowledge for interpreters. Their
participants were improving their language proficiency by spending time abroad
whilst our conscripts spent large amounts of time under formal training in their
native country. The two groups also differ in native language, Swedish or English,
and language acquired, German, Dari, Russian or Arabic. Phonological differences
between native and acquired language is a possible explanation for dissimilar findings.
There is also the very notable difference that our study is based on comparison with a
control group. See the section named ‘Are all results equally valid?’.

Study 4

Diffusion-weighted images were collected in the same cohort as Study 3. The purpose
was evaluation of possible effects of intense language learning on white matter tracts.
By using DTI we have the possibility of investigating the connections between grey
matter areas in the brain in measures of connectivity. Tract Based Spatial Statistics
within the FSL package was used to calculate skeletonized images of the main white
matter tracts in each participant. The resulting images where then analysed in a
similar manner to Study 3, first between groups to estimate changes over time
followed by within group analysis of interpreters to see whether white matter
microstructure before training could predict later language proficiency. The
Proficiency measure is based on the combined ratings of oral and written ability in
the assigned language. In our eyes it represents the learning outcome of the first
semester at the academy. It is of high importance to the students since they need to
pass this mid-year exam to be allowed to continue at the academy. It is well developed
over numerous years. The measure of Struggle is a subjective rating by the head-
teacher at the academy. This measure was designed to coax whether individuals who
had a hard time meeting the requirements of the academy had to work harder;
possibly leading to differential development. This measure suffered from restriction in
range, with no values lower than 6 or higher than 9 on a scale from 1 to 10. This is
illustrated in Figure 3B. As such the measure is weaker then Proficiency, and we
decided against trying to predict rated Struggle with pre-test white matter
microstructure.
Between groups

We did not find any change over time in white matter microstructure when we compared interpreters to controls. This stands in stark contrast to very recent data from Schlegel and colleagues (2012) who find increases in fractional anisotropy in large frontal networks. To evaluate if differences between groups in the two studies could explain the increases (and lack thereof) we compared our well-trained interpreters at pre-test to our university student controls. In the group of interpreters we found lower values of radial diffusivity in areas that are very similar to those found by Schlegel et al. (2012). This is fairly in line with increases in FA since radial diffusivity indicates how much diffusion there is in a radius surrounding a voxel, a decrease can indicate that diffusion is happening along side rather than through an area. See Figure 4A for illustration of the differences between groups at pre-test. Could our interpreter’s white matter be so well developed from earlier language studies that training at the academy failed to influence its structure further?

Predictive value of white matter

If the interpreters, being the select group that they are, had well developed white matter tracts before entry to the academy it should prove predictive of later performance on an individual level. Connections between white matter and experience or performance have been seen before in language (Golestani, Molko, Dehaene, LeBihan, & Pallier, 2007) as well as other modalities (Bengtsson, et al., 2005). We regressed individual white matter values from pre-test measures within the group of interpreters against Proficiency to see if differences in white matter integrity could predict later language performance. That is, can white matter microstructure as it stood prior to training at the academy, help predict how well you will perform as a military interpreter? We found areas that corresponded well to the areas where interpreters as a group showed lower RD values, which in turn corresponds well to the regions reported by Schlegel et al. (2012) but in this case interpreters that performed better following training started their education with higher values of mean diffusivity. In effect, higher values of mean diffusivity in areas connecting different language areas of the brain predicted later language Proficiency. This is counterintuitive since high MD is generally interpreted as a lack of highly directional fibres, making it easier for water molecules to diffuse. A clear glass of water for example, would have very high MD values, whilst water running within a hose would have very high FA. If anything one would expect the opposite, that interpreters with high directionality in fibres would perform better during their stay at the academy. For illustration, see Figure 4 and 5. It is worth noting that all of the areas that increased in grey matter in Study 3 are adjacent to the presented white matter tracts.
Figure 4. A. Interpreters had relatively higher values of radial diffusivity as compared to controls. 4B, C and D; Interpreters with higher Proficiency at the end of 3 months of language training had higher values of mean diffusivity from the start.

So, compared to controls interpreters as a group exhibit lower values of radial diffusivity in tracts that are adjacent to language related areas. This makes sense and corresponds well enough to the findings by Schlegel et al. (2012). On the other hand, interpreters with higher proficiency showed larger values of mean diffusivity in overlapping (and even more widespread) areas. This non-linear relationship shows one tendency of language specialization in the group as a whole, but with relatively less specialized white matter in interpreters that acquired higher Proficiency.
Another recently published study showed similarly confusing results in a group of specialists. In this case, singers were first compared to controls and exhibited lower FA values in tracts of importance. Within the group of singers the inverse was true, more experienced singers showed lower FA values compared to less experienced ones (Halwani, Loui, Rüber, & Schlaug, 2011). Interpreting DTI data is complicated and crossing fibres can lead to negative FA values even if there is a very likely improvement in connection between areas for technical reasons (Jbabdi, Behrens, & Smith, 2010). However, increases in MD does not equal decreases in FA, even if the two measures relate to one another. The effects of crossing fibres on MD was recently investigated by Vos, Jones, and colleagues (2011); MD was demonstrated to be lower in complex white matter configurations (which corresponds well to the findings by Jbabdi, Behrens, & Smith (2010) above) and higher in tissue with a single dominating fibre orientation. It should be noted that the presented TBSS data is contained within the larger tracts of the brain since smaller connections will be averaged out across participants. Hence, high values MD are not an odd finding in itself. Another indicator of crossing fibres is increases or decreases of RD and AD in the same direction, something that is counterintuitive in other instances (since if radial diffusivity is high, with greater diffusion along a tract, axial diffusivity around the tract should be relatively lower). In Paper 4 we find this effect in frontal areas that overlap with the MD effect, both AD and RD also show positive correlations to Proficiency. A possible explanation is that less skilled interpreters had more crossing fibres and higher complexity in the presented tracts; the opposite finding of Halwani et al. (2011). Since we are measuring microscopic effects of myelin, cell membranes and other miniscule structures on diffusion in the scale of millimetres it is outside the scope of this study to delve further into the results.
Figure 5 depicts individual mean values of the voxels within the skeletonized areas in Figure 4 B-D. Mean diffusivity values have been multiplied by 100 for visualization purposes.
Discussion

The overarching goal of the four studies presented in this dissertation was to observe the effects of language acquisition in adults. Earlier findings pointed to grey matter plasticity in the left inferior frontal gyrus following language learning (Stein, et al., 2010). These particular results were partially validated. The inferior frontal gyrus in the left hemisphere did change over time for interpreters, but unlike Stein et al. (2010) it held little importance for measured language proficiency. The studies are highly dissimilar however, in learning situation, group, and methodology. We also knew that white matter networks adjacent to known language areas could change following novel learning of Chinese (Schlegel, et al., 2012). The observed grey matter changes in interpreters are located in these known language areas. A possible explanation why no white matter increases over time were observed in our interpreters may be differences between the groups at start of training. In one case you have a normal student population learning Chinese. Given the setting it is likely that many or some of them were monolingual. The interpreters on the other hand are some of the most gifted multi-linguals that Sweden has to offer a given year. Pre-test differences between interpreters and university student controls mimic the areas found by Schlegel, Rudelson and Tse (2012).

The observed group studied very hard. They had no prior knowledge of their assigned languages. The assigned languages of Arabic, Dari, Russian or Persian were highly dissimilar to their native Swedish. Over the course of one year these individuals went from no prior knowledge to what is likely to be near fluency. Based on the wealth of findings from other training or experience-based studies we wanted to investigate which areas of the cortex would change following language learning. Certainly the inferior frontal gyrus could not be the only area responding to language acquisition? We also expected hippocampal change as well as increases in associative memory performance.

So what did we find? Associative memory-performance increased. The next time your average interpreter attends a party they might be lucky enough to remember 28% more names. We also saw that FreeSurfer can be used with good results in young adults when reliable measurement of the hippocampus is required. As mentioned white matter microstructure did not change as effect of language learning. Perhaps because interpreters already had well developed language networks, being the hard working multi-linguals they are. White matter microstructure did prove predictive of
later language proficiency, with higher values of mean diffusivity in interpreters with good language performance. This unexpected finding might be due to crossing fibres. Although a sea lamprey will probably tell you that it is because of the benefits of having really big axons (Takahashi, et al., 2002).

Our main findings are in Study 3 (Mårtensson, et al., 2012). We found increases in grey matter volume in four areas: the inferior frontal gyrus (Broca’s area), the superior temporal area (Wernicke’s area), the middle frontal gyrus, and the right hippocampus. The right hippocampus increased by 2.55%. This is in line with earlier findings in humans (Draganski, et al., 2006). The left hippocampus, for comparison, showed an insignificant increase of 1%. Cortical areas increased 3% compared to 1-9% in animal studies (Anderson, 2011). These findings correspond very well to two important models of language: Hickok & Poeppel’s dual-stream model (2007) and Davis and Gaskell’s (2009) complementary systems account of word learning. The superior temporal gyrus and the adjacent superior temporal sulcus hold special importance in the dual-stream model since it is suggested as shared tissue between the ventral and dorsal streams that handle perception and production of speech. Interpreters as a group saw increases in the superior temporal gyrus of the left hemisphere. More skilled interpreters showed relatively larger increases in grey matter volume in this key area compared to their less proficient counterparts. Hippocampal increases are in line with Davis and Gaskell (2009) model of word learning, where the hippocampus is predicted to hold importance in rapid word learning and interfaces with other language areas such as the superior temporal gyrus. The presented findings support this notion, which is further strengthened by positive correlations between language proficiency and a tendency towards negative correlation with ratings of effort needed to stay at the academy. The right hippocampus is the only area that exhibits this dissociation between Proficiency and Struggle. The inferior frontal gyrus and middle frontal gyrus are both part of the articulatory network (Davis & Gaskell, 2009). The middle frontal gyrus was the only area that showed larger increases of cortical thickness in participants with higher ratings of Struggle to stay at the academy. A possible explanation is that these participants had a harder time articulating and/or understanding foreign speech (Meister, et al., 2007) and because of this had to direct more time and effort into processes related to articulation and speech perception. One could argue that individual strategies used during learning have influenced which areas developed, with more efficient strategies leading to development in areas that are relatively more important for educational outcomes. At the start of the stay at the academy all interpreters are taught appropriate study technique. Whilst this does not rule out the possibility of students using different strategies for learning new material, it does make this explanation less likely.

Taken as a whole the increases in grey matter volume and their behavioural correlates can be seen as indication that plasticity in the hippocampus and superior temporal
gyrus are especially important for keeping up with the high rate of learning at the academy.

Will language training change local brain volume? Based on our solid findings from Study 3 it will. Is it safe to relax in the knowledge that you won't have to learn how to juggle? Well not yet, remember that this group is hand picked from a large group of language talents and study most of their waking hours. And yet within this group of pre-selected top-students we still saw local increases in brain volume. Even a well-tempered brain develops under strain.

Possible implications for the aging brain

We now know that grey and white matter can change in young adults, hopefully it is possible to achieve similar effects in older adults. Our natural aging processes have a negative effect on several cognitive abilities such as memory and spatial navigation (Balota, Dolan, & Duchek, 2000; Schaie, 1996). Different training forms have been used to induce positive change in older brains, such as: aerobic training (Colcombe, et al., 2006; Erickson, et al., 2011), juggling (Boyke, et al., 2008) and different forms of mnemonic training (Engvig, et al., 2010, 2011; Lövdén, Bodammer, et al., 2010). Importantly, bilingualism has recently been linked to delay of both Alzheimer's disease (Craik, Bialystok, & Freedman, 2010) and dementia (Bialystok, Craik, & Freedman, 2007), which lends a sliver of hope that language training could serve as a possible counter-measure for a deteriorating brain. It should be noted however, that our findings in young healthy adults acquiring a foreign language at record speed far from equates to life-long experience as a bilingual.

Limitations

The presented work has several strengths. The data showing increases in grey matter volume are longitudinal and have survived direct comparison to controls that were matched for age, level of education and measures of intelligence and mental stability. The method used to obtain these results has been validated, both by our own studies and that of others. But our findings are not without limitations. A quasi experimental design was unavoidable due to the setting that the interpreters are in. It would have been practically impossible to gather participants trying to enter the academy, randomising them into groups and comparing the effects of education versus a waiting list or other activity since the academy always wants the very best students, and all of them, each year. The intensity and duration of their training would be very
hard if not impossible to match outside of their context; as such our control groups of university students are imperfect. Another limitation is that of generalization; the interpreters do not in any way represent the general population in Sweden; they are exceptional individuals in a rare situation and as such we cannot draw direct conclusions from their experience to that of the population as a whole. Whilst your average university student is expected to study 40 hours per week interpreters at the academy study most of their waking hours and maintain a high study pace for months at end. More studies into the effects of language training, in lower intensity and in randomized groups, is needed to evaluate whether our findings can be replicated in a less intensive study environment.

Outlook

So where do we go from here? We know that grey matter volume can change in young adulthood given sufficient intensity of training in a foreign language. But we do not know what has actually happened. As mentioned in the beginning of this dissertation grey matter changes can be an indication of several processes; neurogenesis (birth of new neurons); synaptogenesis (creation of new synapses) or change in size. In humans we currently lack the resolution to measure these effects in vivo but in our not too distant relatives (mice) spatial navigation was found to increase hippocampal volume. In this case the increase correlated selectively with markers of neuronal process remodelling rather than neurogenesis, amount of neurons or astrocytes (Lerch et al., 2011). Combined research using both animals and humans such as Pereira et al. (2007) or improved neuroimaging methods can help us understand the structural underpinnings of changes in humans.

Another important focus should be time. Our structural scans have a temporal resolution of 2 pictures in 3 months. Little is known when changes occurred during that time period, or if changes came and left during the course of learning. For comparison Sagi et al. (2012) saw changes after as little as 2 hours of video gaming. Schlegel et al. (2012) scanned 8 times over the course of 9 months. Future studies should focus on increasing temporal resolution. This along with increased spatial resolution as argued above would increase our chances of not only seeing that something has happened, but also what happened and when it happened.

I will leave you with the Zatorre, Fields and Johansen-Berg (2012; page 530) quote that helped outline the purpose of this dissertation. I believe we have done what we set out to do:

There are several reasons why it is important to establish relationships between MRI-based effects and behavior. First, in the context of learning, many different processes may be operating in parallel (for example, learning words of a new language may entail auditory discrimination, motor articulatory skills and semantic memory). If multiple
brain regions show changes, interpretation is enhanced if they can be linked to separable behavioural effects. Second, individual differences in what is learned or how it is learned may not be understood without measuring behavioural outcomes. Finally, for such research to have clinical relevance, it is essential to establish correlations between behaviour and structure if one wants to understand disorders that are diagnosed on the basis of behavioural disturbances.


<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Intervention</th>
<th>Sample</th>
<th>Analysis technique</th>
<th>Statistical technique</th>
<th>Increase</th>
<th>Decrease</th>
<th>Training</th>
<th>Duration</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draganski et al. (2004)</td>
<td>Between subjects</td>
<td>Juggling</td>
<td>12</td>
<td>Exp.</td>
<td>VBM</td>
<td>5</td>
<td>Decrease</td>
<td>Training</td>
<td>3 months</td>
<td>-</td>
</tr>
<tr>
<td>Colcombe et al. (2006)</td>
<td>Between subjects</td>
<td>Aerobics</td>
<td>30</td>
<td>Cont.</td>
<td>VBM</td>
<td>5</td>
<td>-</td>
<td>Increase</td>
<td>3 months</td>
<td>-</td>
</tr>
<tr>
<td>Draganski et al. (2006)</td>
<td>Between subjects</td>
<td>Studying for an exam</td>
<td>38</td>
<td>Exp.</td>
<td>VBM</td>
<td>5</td>
<td>Decrease</td>
<td>Training</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Draganski et al. (2006)</td>
<td>Between subjects</td>
<td>Juggling</td>
<td>25</td>
<td>Cont.</td>
<td>VBM</td>
<td>5</td>
<td>Decrease</td>
<td>Training</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Boyle et al. (2008)</td>
<td>Between subjects</td>
<td>Juggling</td>
<td>25</td>
<td>Cont.</td>
<td>VBM</td>
<td>5</td>
<td>Decrease</td>
<td>Training</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Drumester et al. (2008)</td>
<td>Within subjects</td>
<td>Juggling</td>
<td>25</td>
<td>Cont.</td>
<td>VBM</td>
<td>5</td>
<td>Decrease</td>
<td>Training</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Colcombe et al. (2006)</td>
<td>Between subjects</td>
<td>Reading</td>
<td>25</td>
<td>Cont.</td>
<td>VBM</td>
<td>5</td>
<td>Decrease</td>
<td>Training</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ilg et al. (2008)</td>
<td>Between subjects</td>
<td>Pain stimulation</td>
<td>14</td>
<td>Cont.</td>
<td>VBM</td>
<td>5</td>
<td>Decrease</td>
<td>Training</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Teusch et al. (2008)</td>
<td>Between subjects</td>
<td>Cognitive learning</td>
<td>13</td>
<td>Cont.</td>
<td>TBM</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>2 weeks</td>
<td>20m per day</td>
</tr>
<tr>
<td>Ceccarelli et al. (2009)</td>
<td>Between subjects</td>
<td>Tennis</td>
<td>13</td>
<td>Cont.</td>
<td>CIVET</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>2 weeks</td>
<td>1.5h per week</td>
</tr>
<tr>
<td>Haer et al. (2009)</td>
<td>Between subjects</td>
<td>Pain</td>
<td>15</td>
<td>Cont.</td>
<td>L. BA 6 and, 22/38</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>2 weeks</td>
<td>22/38</td>
</tr>
</tbody>
</table>

Exp. = Experimental, Cont. = Control, M. Age = Mean Age, ROI = Region of Interest, VBM = Voxel-based morphometry, TBM = Tract-based morphometry, CIVET = Connection-based Imaging and Virtual Environment Technology.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Group Design</th>
<th>Task</th>
<th>Duration</th>
<th>ROI Description</th>
<th>FA/MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scholz et al. (2009)</td>
<td>Between subjects</td>
<td>Juggling</td>
<td>24</td>
<td>Medial Occipital and Parietal lobes bilaterally</td>
<td>-</td>
</tr>
<tr>
<td>Thomas et al. (2009)</td>
<td>Within subjects</td>
<td>Visuo-motor</td>
<td>12</td>
<td>Within group with control condition</td>
<td>-</td>
</tr>
<tr>
<td>Engvig et al. (2010)</td>
<td>Between subjects</td>
<td>Mnemonic training</td>
<td>22</td>
<td>R. Insula, Lat. Orbitofrontal and Fusiform cortex</td>
<td>-</td>
</tr>
<tr>
<td>Kim et al. (2010)</td>
<td>Within subjects</td>
<td>Early postpartum</td>
<td>19</td>
<td>Prefrontal cortex, Parietal lobes &amp; Midbrain</td>
<td>-</td>
</tr>
<tr>
<td>Lövdén et al. (2010)</td>
<td>Between subjects</td>
<td>Mnemonic training</td>
<td>12y/20o</td>
<td>Anterior corpus callosum: FA</td>
<td>-</td>
</tr>
<tr>
<td>Schmidt-Wilcke et al. (2010)</td>
<td>Between subjects</td>
<td>Deciphering morse code</td>
<td>16</td>
<td>Left Occipotemporal region</td>
<td>-</td>
</tr>
<tr>
<td>Stein et al. (2010)</td>
<td>Within subjects</td>
<td>Language training</td>
<td>10</td>
<td>L. Inferior frontal gyrus with proficiency added as covariate</td>
<td>-</td>
</tr>
<tr>
<td>Takeuchi et al. (2010)</td>
<td>Between subjects</td>
<td>Working memory</td>
<td>18</td>
<td>DLPFC, Parietal cortices, L. Superior temporal gyrus</td>
<td>-</td>
</tr>
<tr>
<td>Tang et al. (2010)</td>
<td>Between subjects</td>
<td>Meditation</td>
<td>22</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Taubert et al. (2010)</td>
<td>Between subjects</td>
<td>Balancing</td>
<td>14</td>
<td>R. mid. temporal gyrus &amp; Lobule VIII</td>
<td>-</td>
</tr>
<tr>
<td>Bezzola et al. (2011)</td>
<td>Between subjects</td>
<td>Golf</td>
<td>12</td>
<td>C. Sulcus, Premotor cortex, Inf. Par. Lobule, Interpar. sulcus &amp; Parieto-occipital junctions</td>
<td>-</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Intervention</td>
<td>Sample</td>
<td>Analysis technique</td>
<td>Statistical technique</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------------</td>
<td>-----------------------</td>
<td>--------</td>
<td>--------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Erickson et al. (2011)</td>
<td>Between subjects</td>
<td>Aerobics</td>
<td>60 Exp. 60 Cont. 67 M. Age</td>
<td>FSL-FIRST</td>
<td>Groups over time</td>
</tr>
<tr>
<td>Engvig et al. (2011)</td>
<td>Between subjects</td>
<td>Mnemonic training</td>
<td>21 Exp. 20 Cont. 61 M. Age</td>
<td>TBSS</td>
<td>Groups over time</td>
</tr>
<tr>
<td>Granert et al. (2011)</td>
<td>Within subjects</td>
<td>Motor</td>
<td>14 Exp. - Cont. 51 M. Age</td>
<td>VBM</td>
<td>Within group with ROI</td>
</tr>
<tr>
<td>Hamzei et al. (2011)</td>
<td>Within subjects</td>
<td>Motor task</td>
<td>16 Exp. - Cont. 24 M. Age</td>
<td>VBM</td>
<td>Within group</td>
</tr>
<tr>
<td>Hölzel et al. (2011)</td>
<td>Between subjects</td>
<td>Meditation</td>
<td>16 Exp. 17 Cont. 39 M. Age</td>
<td>VBM</td>
<td>Within training group; groups over time within ROI</td>
</tr>
<tr>
<td>Kwok et al. (2011)</td>
<td>Within subjects</td>
<td>Learning color names</td>
<td>19 Exp. - Cont. 20 M. Age</td>
<td>VBM</td>
<td>Within group</td>
</tr>
<tr>
<td>Landi et al. (2011)</td>
<td>Within subjects</td>
<td>Visuo-motor</td>
<td>12 Exp. - Cont. 26 M. Age</td>
<td>VBM</td>
<td>Within group based on ROI from fMRI</td>
</tr>
<tr>
<td>Wenger et al. (2011)</td>
<td>Between subjects</td>
<td>Spatial navigation</td>
<td>23y/23o Exp. 21y/24o Cont. 26y/65o M. Age</td>
<td>FreeSurfer cortical</td>
<td>Each group over time; groups over time within ROI</td>
</tr>
<tr>
<td>Woolett &amp; Maguire (2011)</td>
<td>Between subjects</td>
<td>Spatial memory</td>
<td>39 Exp. 31 Cont. 38 M. Age</td>
<td>VBM</td>
<td>Within each group separately over time</td>
</tr>
<tr>
<td>Gryga et al. (2012)</td>
<td>Within subjects</td>
<td>Motor task</td>
<td>15 Exp. - Cont. 22-32 (range) M. Age</td>
<td>VBM</td>
<td>Within group</td>
</tr>
<tr>
<td>Langer et al. (2012)</td>
<td>Within group</td>
<td>Motor task</td>
<td>10 Exp. - Cont. 32 M. Age</td>
<td>Freesurfer cortical</td>
<td>Within group</td>
</tr>
<tr>
<td>Design</td>
<td>Intervention</td>
<td>Analysis technique</td>
<td>Increase/Decrease</td>
<td>Training duration</td>
<td>Amount</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>-------------------------</td>
<td>----------------------------------------</td>
<td>-------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Between subjects</td>
<td>Between groups</td>
<td>None, effects driven by</td>
<td>None for training g. Ctrl. group declined in HC values</td>
<td>4 months</td>
<td>50m per day</td>
</tr>
<tr>
<td>Lövén et al. (2012)</td>
<td>Spatial navigation</td>
<td>Manual HC tracing</td>
<td>Between groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mårtensson et al. (2012)</td>
<td>Language training</td>
<td>FreeSurfer cort. &amp; subc.</td>
<td>Groups over time</td>
<td>-</td>
<td>3 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sagi et al. (2012)</td>
<td>Car game</td>
<td>VBA</td>
<td>Within group; followed by interaction analysis</td>
<td>-</td>
<td>2hr</td>
</tr>
<tr>
<td>Schlegel et al. (2012)</td>
<td>Language training</td>
<td>FSL &amp; FreeSurfer</td>
<td>Between groups within predefined tracts</td>
<td>Widespread increases in FA in language and frontal areas</td>
<td>Decrease in RD in frontal areas</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 explained. Outline of longitudinal studies of experience-induced grey- and white-matter changes. Studies are organized by publication year followed by alphabetical order within a given year. **Design** indicates whether the study design included use of one or more control groups. **Intervention** corresponds to the main training form used. **Analysis technique** describes the software and method used to pre-process and analyse volumetric data: VBM = Voxel-based Morphometry, TBM = Tensor-based Morphometry. CIVET is used for processing of cortical data. FSL-FIRST is used for subcortical data. FreeSurfer processes cortical and subcortical data in separate processing streams. Please see details in the listed publications for specifics. **Statistical technique** gives a simplified description of pre-processed data was analysed: ‘Groups over time’ is either a repeated measures ANOVA with two factors (group x time) or methods considered equivalent (such as a difference image calculated for each group and then t-tested for differences). ‘Within group’ analysis generally consists of a t-test for time effects but without a control group. The in-between ‘Within group’ followed by x represents analysis within the training group followed by different forms of comparisons with a control group in the areas that were found. **Increase/Decrease**: Changes over time for the training group only, Lövén et al., 2012 being the exception since the interaction effect was fuelled by decline in controls rather than increases in the training group. **Training duration and amount** is noted as mean values when specified. Unless specified by left (L) or right (R) all effects are bilateral. HC = Hippocampus. DLPFC = Dorsolateral prefrontal cortex. Young (y) and Old (o) are specified in studies containing multiple age groups.