THE NEUROCOGNITIVE FINGERPRINT OF SIMULTANEOUS INTERPRETATION

A Dissertation submitted

for the degree of Philosophiæ Doctor

in Cognitive Neuroscience

Laura Elizabeth Babcock

Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy

March 2015

“To have another language is to possess a second soul.”

–Charlemagne

TABLE OF CONTENTS

ABSTRACT

The human brain is a dynamic organ which is molded throughout the lifespan by each individual's life experiences. Learning to juggle, speaking two languages, and playing the piano all number among experiences that leave an impression on the mind and brain. Little research, however, has examined one of the most demanding processes in human cognition, simultaneous interpretation (SI). In SI an individual must comprehend a stream of auditory material in one language and with a few seconds delay produce the same content in another language. This process, which is both a specialized form of bilingualism and a learned skill, similar to playing the piano, likely leaves its own distinct fingerprint on the mind and brain. The present work examines the neurocognition of professional and trainee simultaneous interpreters to better understand the process of simultaneous interpretation and the lasting impression it leaves.

Bilingualism has been previously associated with advantages in cognitive control in both linguistic and non-linguistic domains. These benefits are posited to be due to bilinguals' extensive practice managing two languages. Simultaneous interpretation represents a process which requires a higher level of language management than most bilingual contexts. This increased experience may lead to quantitatively larger benefits in interpreters than in bilinguals. Additionally, interpreters may garner benefits which are unique to the interpreting experience, in particular the need to use two languages simultaneously. The first study addressed these possibilities in an examination of professional interpreters and matched multilinguals on three tasks of cognitive control. The two groups showed no

differences on the color-word Stroop and Attention Network Test, tasks which have previously revealed an advantage for bilinguals. Results from a non-linguistic task-switching paradigm were mixed. Interpreters showed no additional advantage in switching costs, where bilingual benefits have previously been seen, but exhibited smaller mixing costs than the multilinguals. In comparison with previous literature, this benefit in mixing cost appears to be unique to simultaneous interpretation. Additionally, the interpreters had larger verbal and spatial memory spans than the multilinguals. The results suggest that professional interpreters do not have quantitatively larger bilingual benefits, but do possess benefits specific to experience with simultaneous interpretation

As simultaneous interpretation is an acquired skill, these interpreter-specific advantages may have been gained through training in SI or represent innate differences that led individuals to the field. The second study examined students earning a Master in Conference Interpreting, and matched students in other disciplines, longitudinally to determine which cognitive abilities are innate and which are acquired through SI training. The results indicated improvements in verbal and spatial memory among the students of interpretation, but not among the students in other disciplines, suggesting that these abilities are acquired with training. An improvement in the mixing cost, however, was seen across the groups, leaving open the possibility of an influence of professional experience on this measure.

Previous studies of skill acquisition have indicated that not only cognitive abilities, but also underlying brain structure is altered through the training period. To examine the effects of training in simultaneous interpretation on gray matter and white matter structure the above-mentioned groups additionally participated in neuroimaging sessions. Analysis of gray matter volume using voxel-based morphometry (VBM) revealed group differences in regions previously linked to spoken word learning, suggesting greater efficiency in these areas among students of interpretation. Additionally, changes in gray matter volume related to training in SI were evident in bilateral putamen and left superior temporal cortex, among other regions. Previous functional MRI studies of speech shadowing have found activation in these same regions, suggesting the changes may be related to the simultaneity of input and output during simultaneous interpretation. Moreover, analyses of diffusion tensor imaging (DTI) data revealed greater white matter integrity among the students of interpretation in tracts in the left hemisphere that underlie language. A subset of these tracts was further strengthened through training in SI.

Finally, the mechanisms supporting the simultaneous use of two languages were considered, specifically addressing the possibility that interpreters apply less inhibition to the unused language. Students at various stages in their simultaneous interpretation training were tested on a three language switching paradigm. This paradigm affords a measure of inhibition of abandoned task sets through n-2 repetition costs. Though differences were found between the groups on n-2 repetition costs, these did not appear to be connected to SI training, but rather the predominant bilingual interactional context of the groups.

Taken together these investigations begin to provide a picture of the effects that simultaneous interpretation has on cognitive abilities and brain structure. Specifically, interpreters appear to have a unique set of cognitive advantages that are related to the processes used during SI. Further, these advantages originate from a combination of innate and trained abilities.

CHAPTER 1

GENERAL INTRODUCTION

1.1 The Impressionable Brain

The human mind and brain are not static entities, but rather are sculpted by the numerous life experiences of each individual. Learning to juggle, playing a musical instrument, and playing computer games are all known to leave an impression on cognitive abilities and brain structure (e.g., Bialystok & Depape, 2009; Bialystok, 2006; Draganski & May, 2008; Schellenberg, 2005). Perhaps the most well-known (among the public) life experience to sculpt the mind and brain is bilingualism.

The past two decades have produced a wealth of studies on the cognitive effects of using two languages in daily life. These studies largely suggest that bilingualism is associated with cognitive enhancements beyond the language domain. In particular, bilinguals have shown advantages on tasks that require conflict resolution, attentional control, and shifting between mental sets (e.g., Bialystok, Craik, & Luk, 2008; Costa, Hernández, & Sebastián-Gallés, 2008; Prior & Macwhinney, 2010; but see Paap & Greenberg, 2013). The rationale for these benefits beyond the language domain relies on the widely replicated finding that a bilingual's two languages are activated in parallel (see Kroll, Dussias, Bice, & Perrotti, 2015 for a review). This parallel activation appears to be universal, having been noted in both receptive (e.g., Spivey & Marian, 1999; van Hell & Dijkstra, 2002; van Heuven, Schriefers, Dijkstra, & Hagoort, 2008) and productive language (see Kroll, Bobb, Misra, & Guo, 2008 for a review). To control interference between the languages and produce the intended language, bilinguals are thought to use domain-general cognitive control mechanisms (Bialystok et al., 2008; Green, 1998). Thus, given that humans use language during virtually all waking hours, bilinguals have accumulated immeasurable practice with these cognitive control mechanisms.

1.2 What about Simultaneous Interpretation?

Simultaneous interpretation (SI) is perhaps one of the most cognitively demanding tasks of human cognition. An interpreter must simultaneously (1) comprehend and store a continuous stream of auditory material in one language, (2) reformulate the content of a previous segment into another language, and (3) articulate a previously reformulated segment. Thus, simultaneous interpretation could be considered an extreme form of bilingualism. As for bilinguals, interpreters must manage two linguistic codes and the interference between these codes. Simultaneous interpretation may, however, demand a higher level of control as the interpreter must comprehend one language while producing and monitoring another language. Given this greater need for control, interpreters may accumulate more language management practice than bilinguals and accordingly show a greater advantage in cognitive control. Additionally, simultaneous interpretation involves a host of other processes, such as simultaneous input and output and memory, which may sculpt the mind and brain in areas unique to SI. Despite the potential that simultaneous interpretation has to mold cognition, relatively few studies, especially by neuroscientists, have examined the cognitive changes associated with this complex skill.

1.2.1 Previous Evidence

Among the studies that have investigated cognition in interpreters, three main areas have been considered. First, memory has been tested using tasks of verbal short-term and working memory. The overwhelming result is that professional interpreters have larger verbal working memory and verbal short-term memory spans than various control groups (Bajo, Padilla, & Padilla, 2000; Christoffels, de Groot, & Kroll, 2006; P. Padilla, Bajo, Cañas, & Padilla, 1995; Signorelli, Haarmann, & Obler, 2011; Stavrakaki, Megari, Kosmidis, Apostolidou, & Takou, 2012; Yudes, Macizo, & Bajo, 2011, 2012; but see Köpke & Nespoulous, 2006). Related to memory, the second commonly studied area of cognition is articulatory suppression. This is the process of blocking rehearsal of information in the phonological loop of working memory by repetition of unrelated speech during a memorization task. The normal finding is that recall is hindered by articulatory suppression, however, interpreters have been found to be less affected by articulatory suppression (Bajo et al., 2000; F. Padilla, Bajo, & Macizo, 2005; P. Padilla et al., 1995; Yudes et al., 2012; but see Köpke & Nespoulous, 2006). Finally, a few studies have examined measures of lexical access among professional interpreters and found faster responses in picture naming, lexical decision, and lexical categorization (Bajo et al., 2000; De Groot & Christoffels, 2006), as well as higher verbal fluency[1] (Stavrakaki et al., 2012).

These three areas, which show advantages specific to interpreters, are closely linked to the processes required by simultaneous interpretation. Verbal memory is recruited to store the input language and reformulated output. Articulatory

[1] It is interesting to note that lexical access is an area of cognition where bilinguals typically show worse performance than monolinguals (Michael & Gollan, 2005). Thus it would be interesting to compare the performance of interpreters and monolinguals on tasks of lexical access to determine if interpreters have merely recovered this deficit or improved even further.

suppression is a mirror of the SI process as content must be stored while other material is articulated. Finally, faster retrieval of lexical items should decrease the processing load during SI. Thus, the examination of these areas as well as the result of enhanced abilities among interpreters is unsurprising.

Fewer studies, however, have examined in professional interpreters the executive control functions which typically show benefits in bilinguals. It is precisely these tasks, however, which will allow us to understand if simultaneous interpreters should be considered expert bilinguals. Of the three studies which examined conflict resolution and attentional control, none found an advantage for interpreters (Köpke & Nespoulous, 2006; Morales, Padilla, Gómez-Ariza, & Bajo, 2015; Yudes et al., 2011).

1.2.2 The Effects of Training

Unlike bilingualism, simultaneous interpretation is a necessarily acquired skill. Thus we may ask whether the advantages seen in interpreters are a result of training in SI or rather are innate abilities that contributed to selection and success in the field. Only two studies have examined the longitudinal changes in cognition associated with learning the simultaneous interpretation skill. Bajo and colleagues found that training in simultaneous interpretation was associated with improvements in lexical access (Bajo et al., 2000). A study of students training in American Sign Language (ASL) interpretation, however, showed mixed results with improvements in task-switching, mental flexibility, and working memory tasks involving transformation, but not in working memory tasks involving storage and processing (Macnamara & Conway, 2013). It should be noted, however, that this latter study did not include a control group, so it is unclear if the improvements are due to training or test-retest effects.

1.2.3 How is Simultaneous Interpretation Achieved?

As previously mentioned, SI requires the simultaneous use of two languages. Thus, the language control mechanisms employed in bilingual contexts, which rely on inhibition of the unintended language, may not be of use in simultaneous interpretation. Interpreters may instead maintain both languages active. Preliminary support for this hypothesis is provided by Ibáñez and colleagues (2010) who demonstrated greater cognate facilitation in translators (with at least two years of interpretation experience) than bilinguals.

To comprehend the full picture of simultaneous interpretation three studies have employed neuroimaging techniques (Aherns, Kalderon, Krick, & Reith, 2010; Hervais-Adelman, Moser-Mercer, Michel, & Golestani, 2014; Rinne et al., 2000). Common to all three studies were activations in left motor regions, left inferior frontal gyrus, left inferior temporal gyrus, and right cerebellum, suggesting a primary recruitment of language regions. While neuroimaging provides a unique look at the brain regions involved in simultaneous interpretation, it is not without difficulty. Neuroimaging requires participants to remain motionless during acquisition. Thus to accommodate simultaneous interpretation which requires speech, either the acquisition technique or process must be altered.

1.3 Overview of the Project

The current project aims to extend our knowledge of the complex process of simultaneous interpretation by examining the imprint it leaves on the mind and brain. In the first study we add to the literature on the cognitive advantages of simultaneous interpretation. In particular, we examine whether interpreters should be considered expert bilinguals, who show the benefits of bilingualism above and beyond those seen in matched multilinguals, or if they are unique bilinguals who have a specific set of advantages which are related to the simultaneous interpretation skill. In the second study we aim to determine whether the advantages seen in interpreters are due to innate differences that influenced their selection of interpretation as a career and their ultimate success, or rather are due to abilities that are specifically trained through acquisition of the simultaneous interpretation skill. In the third study we consider the structural correlates of simultaneous interpretation, specifically looking for regions that are changed by training and regions that are predictive of initial success. Finally, we address the role of inhibition in language control during simultaneous interpretation. Using a cross-sectional design we examine how inhibition is applied to language control in students with varying levels of interpretation experience.

CHAPTER 2

ARE INTERPRETERS EXPERT BILINGUALS, UNIQUE BILINGUALS, OR BOTH?

2.1 Introduction

Bilinguals are expert language managers. They navigate the world in two languages, selecting at every moment the appropriate language to use while inhibiting the unintended language. Research over the past two decades has shown that this daily practice in language management leads to benefits in executive functioning beyond the language domain. In particular, advantages have been seen on tasks such as the Simon task, the flanker task, and the task-switching paradigm (e.g., Bialystok, Craik, & Luk, 2008; Costa, Hernández, & Sebastián-Gallés, 2008; Prior & Macwhinney, 2010). Experience with two languages appears then to have conferred upon bilinguals expertise in directing attention, managing competition, and resolving conflict more broadly.

At the extreme end of the bilingual language management spectrum lie simultaneous interpreters. In simultaneous interpretation (SI) an individual attends to a stream of oral material in one language and with a few seconds delay produces the same content in another language. Similar to other bilingual contexts, SI requires the management of two linguistic codes and control between the two languages. However, it uniquely involves the simultaneous comprehension of one language and production of another language. This addition of simultaneity of two language systems mandates an extraordinary level of control between the languages.

Experience with this higher level of control may endow interpreters with an enhancement in language management abilities beyond that due to bilingualism alone. As increased language management has been linked to enhanced non-linguistic executive functioning in bilinguals, the same cognitive benefits may then be seen in interpreters when compared to bilinguals.

Language control during SI may not only be more demanding, but also qualitatively different than during other bilingual contexts. In most bilingual contexts both languages are available (Dijkstra, Grainger, & van Heuven, 1999; Kroll, Bobb, & Wodniecka, 2006; Marian, Spivey, & Hirsch, 2003; Rodriguez-Fornells et al., 2005; van Heuven et al., 2008), but only one language is used at any given moment. To combat interference between the languages, bilinguals are thought to inhibit the unused language (Dijkstra & van Heuven, 1998; Green, 1998). Interpreters, however, must comprehend the input language while monitoring their output in the other language, effectively requiring simultaneous comprehension of both languages. It therefore may not be possible for them to rely solely on inhibition as a method of language management. Instead, interpreters may maintain both languages active. Support for this hypothesis comes from Ibáñez and colleagues (2010). In that study bilinguals and translators (with at least two years of interpretation experience) were asked to read and repeat sentences in Spanish and English which included cognate and matched control words. The translators showed faster processing of the cognate words than the control words in both languages, while the bilinguals showed no difference between cognate and control words. Faster processing of cognates is typically understood to indicate the simultaneous activation of two languages. Thus, the translators, but not the bilinguals, appear to have maintained both languages active. The different skills associated with this different type of language management, in particular the maintenance of two language sets, may lead to enhancements in cognitive control that are unique to simultaneous interpreters.

The present study investigated the cognitive control benefits associated with simultaneous interpretation, considering whether these are an extension of the previously-seen benefits of bilingualism or are specific to SI. To this end, we tested professional simultaneous interpreters and a group of well-matched multilinguals on a battery of tasks focused on executive functioning. We report results from three tasks, which were selected based on having reliably shown bilingual benefits in previous research and having the potential to explore benefits unique to SI.

The Stroop task has been widely used to examine executive control and conflict resolution. In the classic color-word version, participants are asked to indicate the ink color of a written color word; the written word can be either congruent or incongruent with the ink color. The additional cost on incongruent trials, dubbed the Stroop effect, is taken as a gauge of conflict resolution (Stroop, 1935). Previous studies focused on bilingual differences have found a smaller Stroop effect for bilinguals compared to monolinguals (e.g., Bialystok et al., 2008; Hernández, Costa, Fuentes, Vivas, & Sebastián-Gallés, 2010). This advantage has been attributed to bilinguals' practice resolving the conflict that occurs between competing lexical items from bilinguals' two languages.

In the non-linguistic domain, the flanker task has often been employed to explore executive control and conflict resolution. In this paradigm, participants are presented with five arrows and asked to indicate the direction of the central arrow; the flanking arrows point in either the same (congruent trials) or opposite direction (incongruent trials). Bilinguals have shown both a smaller difference between congruent and incongruent trials and overall faster responses on this paradigm than monolinguals (e.g., Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Costa et al., 2008; Luk, De Sa, & Bialystok, 2011). A modified version of the traditional flanker task was included in this study, the Attention Network Test (ANT; Fan, McCandliss, Sommer, Raz, & Posner, 2002). This task allows the examination of the alerting and orienting networks in addition to the executive control network. Unlike the executive control network, bilingual advantages have not typically been seen in the orienting (e.g., Costa et al., 2009, 2008; Hernández et al., 2010) and alerting networks (e.g., Costa et al., 2009; Poarch & van Hell, 2012; but see Costa et al., 2008). These networks may, however, reveal advantages specific to SI. Interpreters often need to predict the coming input and therefore make use of several contextual cues while interpreting (e.g., Seeber & Kerzel, 2011). Additionally, they must maintain a high level of alertness.

The final task considered was a non-linguistic task-switching paradigm. In this paradigm, participants must select a response on each trial based on one of two possible task sets (e.g., color or shape discrimination). They are asked to make judgments in single-task blocks in which the task is the same on every trial and in mixed-task blocks in which the task can be either of the two on any given trial. The mixed-task blocks lead to two types of trials, repetition trials in which the current task is the same as the task on the previous trial and switch trials in which the current task is different than the task on the previous trial. Completion of this paradigm involves the use of two dissociable control processes: transient (or local) control and sustained (or global) control (Braver, Reynolds, & Donaldson, 2003; Koch, Prinz, & Allport, 2005; Kray & Lindenberger, 2000). Transient control is used on a trial-by-trial basis to switch between task sets on the switch trials in the mixed-task blocks. Its specific component processes may include updating of goals, internally reconfiguring the task set, and connecting task cues to the correct stimulus-response mappings (Braver et al., 2003). These are potentially the same processes required to switch between languages. Indeed, bilingual advantages in transient control have been noted previously (Garbin et al., 2010; Prior & Gollan, 2011; Prior & Macwhinney, 2010; but see Hernández, Martin, Barceló, & Costa, 2013; Paap & Greenberg, 2013). In addition to the transient control needed on individual trials, the mixed-task block also requires sustained control processes that include the active maintenance of multiple task sets and attentional monitoring for task changes (Braver et al., 2003). These processes are reminiscent of the increased demands of simultaneous interpretation, in particular, the need to maintain multiple task sets in parallel. Notably, differences in sustained control have not typically been found in bilinguals (Prior & Gollan, 2011; Prior & Macwhinney, 2010).

The selected tasks allow for the examination of both increased bilingual benefits in interpreters and benefits specific to simultaneous interpretation. Differences between the professional interpreters and multilinguals in conflict resolution on the Stroop and ANT tasks and in transient control on the task-switching paradigm would suggest that interpreters' more demanding practice endows them with greater benefits. While any differences between the groups on the ANT in the alerting and orienting networks or on the task-switching paradigm in sustained control would indicate that interpreters develop enhancements that are unique to their experience with SI. Benefits of both types may also be seen as they are not in principle mutually exclusive.

2.2 Methods

2.2.1 Participants

Twenty-three professional interpreters (18 females) and twenty-one multilinguals (17 females) participated in the study. All participants had normal or corrected-to-normal vision and reported no history of neurological or psychological problems. Additionally, participants were required to speak English since that was the language chosen for testing and consent. All but one participant reported normal color vision, which was confirmed with the Ishihara Color Test (Ishihara, 1972). One multilingual participant had a below normal score on the Ishihara Color Test, however his data were not excluded because he was not identified as an outlier from his group on any task. Additionally, his performances on the color and shape tasks in the task-switching paradigm were comparable. All participants were living and working in the Brussels area at the time of testing. The professional interpreters' group was formed of individuals working in simultaneous interpretation at the time of testing with at least one year of experience (mean = 13 years, range 1-42 years) and forty-five working days per year (mean = 149 days, range 45-234 days). The multilingual individuals used at least two languages on a daily basis and reported no experience with simultaneous interpretation. Importantly, the two groups were matched on a number of biographical factors to ensure that any differences seen were due to experience with interpretation and not to other underlying causes. These factors included age, years of education, intelligence (measured with Raven's Advanced Progressive Matrices; Raven, Raven, & Court, 1998), and socioeconomic status (using the proxy of mother's years of education, Gottfried, Gottfried, Bathurst, Guerin, & Parramore, 2003; Noble, McCandliss, & Farah, 2007; Stevens, Lauinger, & Neville, 2009; see Table 1).

The groups were additionally matched on factors relating to their language experience. These data were collected through a language history questionnaire. Participants provided information about all the languages they knew and/or studied. For each language they were asked to detail how and when they learned the language, including immersion experiences, as well as to provide a self-rating in the areas of reading, writing, speaking, and understanding on a 7-point Likert scale. Participants were also asked to evaluate how often they switched between languages within a conversation using a 7-point Likert scale. This was considered in periods throughout their lifetime at home, with friends, and at work (each situation evaluated separately), as well as within the year prior to testing for the following actions: thinking, dreaming, talking to oneself, and expressing anger and affection. Finally, they were asked to quantify using percentages how much they used each language in the above periods and situations and some additional situations evaluated in the year prior to testing.

Based on these data, the two groups were matched on their number of native languages (defined as languages learned and used regularly in the first four years of life) and functional languages at the time of testing (defined as languages receiving an average self-rating across the four areas of at least 4). These factors were important to match as previous studies have shown advantages in executive functioning for early bilinguals compared to late bilinguals (e.g., Bialystok, Craik, & Luk, 2012; Luk et al., 2011; Tao, Marzecová, Taft, Asanowicz, & Wodniecka, 2011) and because it is unclear whether languages beyond the second language contribute to greater benefits (e.g., Bialystok, Craik, Green, & Gollan, 2009; Kavé, Eyal, Shorek, & Cohen-Mansfield, 2008). Additionally, the two groups were matched on their switching frequency in most situations. They differed on only three period-situation pairs: with friends after age 25 (p = .031), with friends in the year prior to testing (p = .001), and marginally at work in the year prior to testing (p = .071). In all three cases, the multilingual group reported a higher rate of switching than the interpreters. Finally, there was no difference in English proficiency between the groups (p = .953), which was on average very high (mean = 6.1, range 4-7).

Table 1: Biographical factors

	Interpreters (N=23)	Multilinguals (N=21)	p-value
Age (in years)	39.3 (13.1)	34.1 (10.3)	p = .149
Years of education	18.2 (1.4)	18.3 (2.6)	p = .913
Raven's APM score	35.0 (4.9)	35.7 (5.6)	p = .695
Mother's years of education	14.2 (3.7)	14.4 (3.6)	p = .884
Number of native languages	1.43 (0.66)	1.43 (0.75)	p = .977
Number of functional languages	4.74 (1.57)	4.14 (1.06)	p = .152
Native languages	Berber, Catalan, Croatian, Dutch, English, French, German, Italian, Polish, Portuguese, Romanian, Spanish	Bulgarian, Czech, English, Filipino, French, Galician, German, Greek, Hiligaynon, Hungarian, Italian, Kinaray-a, Polish, Portuguese, Romanian, Slovak, Spanish, Valencian	

Note: Values reported are means with standard deviations in parentheses.

All participants gave written informed consent and were offered compensation for their time. The study was approved by the the ethical committees of the Scuola Internazionale Superiore di Studi Avanzati (SISSA) and the Faculty of Psychological Sciences and Education at the Université Libre de Bruxelles.

2.2.2 Tasks and Procedure

2.2.2.1 Color-word Stroop task. This task was a translated and shortened version of the task presented in Puccioni and Vallesi (2012). Stimuli consisted of four English color words (BLUE, RED, GREEN, YELLOW) presented individually in ink of one of the four colors (blue, red, green, yellow). English was chosen as the language for this task since it was a guaranteed common language among participants. Additionally, since the two groups did not differ in English proficiency, any potential group differences should not be due to a difference in proficiency. Participants were asked to ignore the word and identify the ink color by pressing the correspondingly colored button on a Cedrus RB-834 response pad (www.cedrus.com; the color-button mappings were counterbalanced across participants). The participants were asked to respond as quickly and accurately as possible. Each stimulus was categorized as congruent (e.g., BLUE presented in blue ink) or incongruent (e.g., BLUE presented in red ink). Roughly half of the trials were congruent and half incongruent. Only complete alternation sequences were employed, meaning that neither the ink nor the word color(s) used in trial *n* were used in either way (ink or word) in trial *n+1*, thus minimizing both positive and negative priming confounds (see Puccioni & Vallesi, 2012 for details).

The task included two blocks of 64 trials each with a short rest break between the blocks. Trials consisted of stimulus presentation in the center of the screen for 500 ms, a 2000 ms blank response screen, and an additional blank screen beyond the response time limit which lasted randomly and continuously between 250 and 700 ms. Prior to the experimental blocks participants completed a training block to ensure that all participants understood the task. This training block was composed of 16 items representing all possible word-ink combinations. Items were presented on screen until a response was made. Feedback about accuracy and speed followed the response and lasted on screen for 1200 ms, followed by a 500 ms inter-trial interval. All participants reached the criterion (10 correct trials out of 16) to move onto the experimental trials in one run of the training block.

The primary comparison to be drawn from this task is the difference in accuracy and response time to congruent and incongruent trials, termed the Stroop effect. This difference gives a measure of conflict resolution.

2.2.2.2 ANT. This task was adapted from Costa and colleagues (2008). The target stimuli consisted of five arrows situated either above or below a central fixation cross. The four "outside" arrows pointed in a uniform direction, while the central arrow could point in either the same direction as the others (congruent) or the opposite direction (incongruent). The congruent condition comprised 75% of the trials and the incongruent condition 25%. Participants were asked to indicate the direction of the central arrow using the leftmost and rightmost buttons on a Cedrus RB-834 response pad. Prior to each target stimulus a cue appeared which belonged to one of four types: no cue, central cue, double cue, and spatial cue. In no cue trials the fixation cross remained throughout the cue period. The central cue was an asterisk in the place of the fixation cross. The double cue was the fixation cross plus asterisks at both potential locations of the central arrow (above and below the fixation cross). The spatial cue was the fixation cross plus an asterisk at the location where the central arrow would occur (either above or below the fixation cross).

The task included two blocks of 128 trials each with a short rest break between the blocks. Each trial began with a 400 ms fixation cross followed by the cue (no cue, central cue, double cue, or spatial cue), which appeared for 100 ms, followed by fixation for another 400 ms. The target then appeared and remained on screen until the end of the trial which was marked by the participant's response or the expiration of 1700 ms. Participants completed 8 practice trials prior to the experimental blocks.

This task allows the examination of three attentional networks devoted to executive function, alerting, and orienting. The executive function network is measured with the conflict effect which is the difference in accuracy or response time between congruent and incongruent trials. The difference between trials with no cue and those with a double cue provides a measure of the alerting network. Finally, the orienting effect is calculated as the difference between trials with a spatial cue and trials with a central cue.

2.2.2.3 Task-switching paradigm. The paradigm was a modified version of the paradigm used in Rubin and Meiran (2005). Stimuli were red and blue hearts and stars presented individually on a white background. On each trial participants were asked to respond to either the color or the shape of the stimulus. The task to be completed was indicated by a visual cue located above the stimulus. To limit the use of linguistic information graphic cues were used. The color task cue consisted of three colored rectangles (purple, orange, and yellow) arranged linearly. Similarly, the shape task cue consisted of three black shapes (triangle, circle, and square) arranged linearly. Participants were required to make a choice response to each trial using the leftmost and rightmost buttons on a Cedrus RB-834 response pad. The four possible response-to-button mappings (left: red/heart, right: blue/star; left: red/star, right: blue/heart; left: blue/heart, right: red/star; left: blue/star, right: red/heart) were counterbalanced across participants.

Trials began with a fixation cross presented for 1500 ms followed by cue presentation. Two cue-to-target intervals (CTI) were employed (100 or 1000 ms), which were distributed randomly and equally across trials. This choice allowed the examination of potential differences in endogenous and exogenous task reconfiguration (Meiran, 1996; Rogers & Monsell, 1995). Following the CTI, the stimulus was presented in the center of the screen, below the cue, which remained onscreen. The trial concluded when the participant gave a response. Incorrect responses were followed by a 100 ms beep.

Participants completed five blocks of trials which formed a sandwich design. Blocks 1, 2, 4, and 5 were single-task blocks in which only one task (color or shape) was presented for the entire block. The same task was presented in blocks 1 and 5 and the other task in blocks 2 and 4; the specific assignment was counterbalanced across participants. The single-task blocks each consisted of 6 practice trials and 24 experimental trials. Block 3 was a mixed-task block with half of the trials requiring a color judgment and the other half a shape judgment. This block included 10 practice trials followed by 192 experimental trials with a short rest break at the halfway point. Half of the trials were repetition trials in which the task to be completed was the same as on the previous trial and half were switch trials in which the task was different than on the previous trial.

The three trial types (switch, repetition, and single-task) lead to two main comparisons. The comparison of the switch and repetition trials in the mixed-task block is informative about the transient control needed to switch tasks. The difference in response time (RT) or accuracy between these trial types is termed the switching cost. Comparing the repetition trials in the mixed-task block and trials in the single-task block provides a gauge of the sustained control needed in the mixed-task block.

This difference in RT or accuracy is referred to as the mixing cost. Previous studies have suggested that the mixing cost is due in part to the increased memory load of the mixed-task block (Meiran, 1996; Rogers & Monsell, 1995). Additionally, several studies have demonstrated that interpreters have larger verbal memory spans than non-interpreter bilinguals (Bajo et al., 2000; Christoffels et al., 2006; Köpke & Nespoulous, 2006; P. Padilla et al., 1995). To disentangle the effects of increased memory span from SI experience on potential group differences, we additionally considered tests of verbal and spatial memory.

2.2.2.4 Memory tests. Short-term memory (STM) was assessed in the verbal and spatial domains using computerized versions of the letter span and matrix span tasks (Kane et al., 2004). These tasks were comparable in their format; participants viewed a sequence of items of variable length and were asked to recall the items in the order they were presented at the end of each sequence. Three sequences of each length were presented with the length selected randomly on each trial. Performance was measured by the number of recalled items in perfectly recalled sequences. The to-be-recalled items in the letter span task were twelve consonants. Each item was presented for 1000 ms and the length of sequences ranged from three to eight items. In the matrix span task the items consisted of a 4x4 grid with one square colored red, the position of the red square was the to-be-recalled item. Each item was presented for 650 ms and sequences contained two to seven items.

Verbal and spatial working memory (WM) were assessed using the automated operation span task and the automated symmetry span task, respectively (Unsworth, Heitz, Schrock, & Engle, 2005). These tasks followed the same format. Similar to the STM tasks, participants were asked to recall sequences of items of varying length, however prior to each item of the sequence an intervening task was presented.

Participants were trained on each task component separately and together before completing the test sequences. In the operation span task the to-be-recalled items were the twelve consonants used in letter span task. The intervening task was an arithmetic operation (e.g., (2x6) - 4 = ?). Sequences consisted of three to seven operation-letter pairs. The to-be-recalled items in the symmetry span task were identical to those in the matrix span task. The intervening task required a symmetry judgment. Sequences contained two to five symmetry-square pairs. In both tasks three sequences of each length were presented with the length randomly selected on each trial. In addition to the number of items in perfectly recalled sequences, the number of errors on the intervening task was also recorded. This included incorrect responses and responses that required a much longer than average response time (calculated during intervening task training).

2.3 Results

2.3.1 Memory Tests

The interpreters performed better on both the verbal and the spatial short-term memory tasks than the multilinguals ($t(42) = 2.1$, $p = .041$, $d = .639$; $t(42) = 3.7$, $p < .001$, $d = 1.133$, respectively; Figure 1). Additionally, the interpreters recalled more items on the verbal working memory task than the multilinguals ($t(42) = 2.1$, $p = .039$, $d = .582$; Figure 1), but the groups did not differ in the number of errors ($p = .329$). There were no differences between the groups on the task of spatial working memory (ps $\geq .299$).

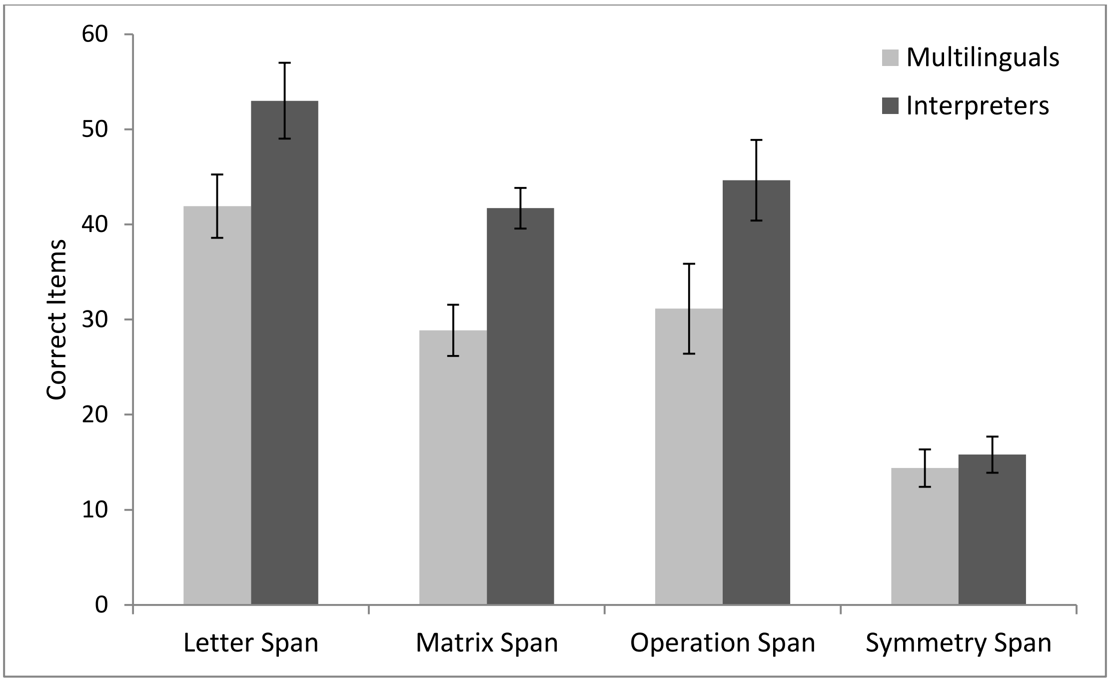

Figure 1: Performance on tests of memory by group. Error bars represent standard errors of the mean. The number of items, and therefore maximum score, for the memory tests were as follows: Letter Span – 99; Matrix Span – 81; Operation Span – 75; Symmetry Span – 42.

2.3.2 Color-Word Stroop Task

Two participants (both male interpreters) were identified as extreme outliers within their group based on their accuracy rate (more than 3 interquartile ranges below the 1st quartile); their data were excluded from all analyses. Their exclusion, however, did not change the matching of the two groups on the abovementioned biographical variables. For all analyses on accuracy the first trial in each block was not considered. Since the accuracy data were non-normally distributed, non-parametric Mann-Whitney U tests were used for their analysis. For the analysis of response time (RT) data, the first trial in each block and error trials were excluded (2.3% of trials). Additionally, for each participant, trials with an RT more than 3 standard deviations (SD) from their individual mean were excluded (0.9% of trials). Finally, trials following an error were excluded to avoid post-error slowing confounds (Burns, 1965; 2.1% of trials).

Analyses on the accuracy data revealed that the groups did not differ on congruent (p = .693) or incongruent trials (p = .230). Further the groups showed no difference on the accuracy Stroop effect (p = .351). A two-way ANOVA on RTs with trial type (congruent, incongruent) and group as factors revealed a main effect of trial type ($F(1,40) = 96.2$, $p < .001$, $\eta_p^2 = .706$). No differences were seen between the groups either overall (p = .345) or in the size of the Stroop effect (p = .570; Table 2).

Table 2: Color-word Stroop: mean response times and accuracy by group

	Interpreters		Multilinguals	
	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)
Congruent trials	626 (110)	97.8 (1.9)	654 (92)	97.9 (1.9)
Incongruent trials	673 (120)	97.9 (2.4)	706 (92)	97.2 (2.3)
Stroop effect	47 (29)	-0.1 (2.7)	53 (36)	0.7 (2.5)

Note: Standard deviations are presented in parentheses.

2.3.3 ANT

One participant (a male interpreter) was identified as an extreme outlier based on his accuracy rate (more than 3 interquartile ranges below the 1st quartile); his data were excluded from all analyses. His exclusion, however, did not change the matching of the two groups on the above-mentioned biographical variables. For all analyses on accuracy, the first trial in each block was not considered. Since the accuracy data were non-normally distributed, non-parametric tests were used for their analysis. Wilcoxon Signed Rank tests were used to compare conditions within each group and Mann-Whitney U tests were used to compare conditions between the two groups. For the analysis of RT data, the first trial in each block and error trials were excluded (1.5% of trials). Additionally, for each participant, trials with an RT more than 3 SDs from their individual mean were excluded (1.4% of trials). Finally, trials following an error were excluded to avoid post-error slowing confounds (Burns, 1965; 1.3% of trials).

2.3.3.1 Conflict effect. The conflict effect analyzes the difference between congruent and incongruent ANT trials. The two groups did not differ in accuracy on either the congruent (p = .256) or incongruent trials (p = .834). Further, there was no difference between the groups on the accuracy conflict effect (p = .535), though both groups showed a reliable difference between congruent and incongruent trials (ps ≤ .002). A two-way ANOVA on RTs with trial type (congruent, incongruent) and group as factors revealed a main effect for trial type ($F(1,41) = 276.0$, $p < .001$, $\eta_p^2 = .871$). The interpreters showed marginally faster RTs ($F(1,41) = 3.2$, $p = .078$, $\eta_p^2 = .074$; Table 3), but did not differ from the multilinguals in the size of the conflict effect (p = .170).

Table 3: ANT: mean response times and accuracy by group

	Interpreters		Multilinguals	
	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)
Congruent trials	444 (55)	99.1 (1.3)	470 (55)	99.1 (0.9)
Incongruent trials	520 (61)	96.4 (3.6)	559 (73)	96.7 (3.0)
Conflict effect	75 (23)	2.8 (2.7)	89 (40)	2.4 (2.6)
Alerting effect	15 (16)	-0.4 (2.1)	11 (14)	-1.1 (2.1)
Orienting effect	4 (18)	0.9 (2.1)	10 (23)	1.0 (2.9)

Note: Standard deviations are presented in parentheses.

2.3.3.2 Alerting effect. The alerting effect was defined as the difference between ANT trials cued with the double cue and those with no cue. The two groups showed no difference in the accuracy alerting effect (p = .296). Additionally, there was no difference between the groups in the RT alerting effect (*p* = .401), though a reliable difference between no cue and double cue trials was seen across the groups ($t(42) = 5.5$, $p < .001$, $d = .847$).

2.3.3.3 Orienting effect. The orienting effect was defined as the difference between central cue trials and spatial cue trials in the ANT. The two groups showed no difference in the accuracy orienting effect (*p* = .902). Additionally, there was no difference between the groups in the RT alerting effect (*p* = .397), though a reliable difference between spatially and centrally cued trials was seen across the groups ($t(42) = 2.2$, $p = .031$, $d = .341$).

2.3.4 Task-Switching Paradigm

Two participants (one male interpreter and one female multilingual) were identified as extreme outliers within their group based on their accuracy rate (more than 3 interquartile ranges below the 1st quartile); their data were excluded from all analyses. Their exclusion, however, did not change the matching of the two groups on the above-mentioned biographical variables. For all analyses on accuracy the first trial in each block and the first trial after the rest break in block 3 were not considered. Since the accuracy data were non-normally distributed, non-parametric tests were used for their analysis. Wilcoxon Signed Rank tests were used to compare conditions within each group and Mann-Whitney U tests were used to compare conditions between the two groups. For the analysis of response time data the first trial in each block, the first trial after the rest break in block 3, and error trials were excluded (3.2% of trials). Additionally, for each participant, trials with an RT more than 3 SDs from their block-type mean (single-task, mixed-task) were excluded (1.6% of trials). Finally, trials following an error were excluded to avoid post-error slowing confounds

(Burns, 1965) and since these trials are likely to be mis-categorized as switch or repetition trials (2.1% of trials).

Before analyzing the switching and mixing costs, the color and shape tasks were compared. Accuracy rates were comparable for the two tasks within each group (*p*s > .122). A two-way ANOVA on RTs revealed no main effect of task (color, shape) and no interaction with group (*p*s > .220). Therefore the data for the two tasks were collapsed for all further analyses.

2.3.4.1 Switching costs. Switching costs represent the difference in performance on switch trials and repetition trials in the mixed-task block. The two groups did not differ in accuracy on either the switch (p = .521) or the repetition trials (p = .940). Additionally, there was no difference between the groups on the accuracy switching cost (p = .345), though both groups showed a reliable cost (*p*s ≤ .028; Table 4). A three-way ANOVA on RTs with CTI length and trial type (switch, repetition) as within-subjects factors and group as a between-subjects factor showed main effects for all three factors. Responses were faster on long compared to short CTI trials ($F(1,40) = 358.9$, $p < .001$, $\eta_p^2 = .900$) and on repetition compared to switch trials ($F(1,40) = 75.9$, $p < .001$, $\eta_p^2 = .655$). Additionally, these factors interacted, revealing a smaller switching cost on long CTI trials than on short CTI trials ($F(1,40) = 32.5$, $p < .001$, $\eta_p^2 = .449$). Interpreters were overall faster than multilinguals ($F(1,40) = 5.6$, $p = .022$, $\eta_p^2 = .124$). The group factor did not interact with either CTI length (p = .299) or trial type (p = .411) suggesting that interpreters and multilinguals did not differ in switching cost. The three-way interaction was also not significant (p = .526).

Table 4: Task-switching paradigm: mean response times and accuracy rates by group

	Interpreters		Multilinguals	
	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)
Trial Type				
Single-task	460 (77)	98.5 (2.2)	519 (58)	98.9 (1.3)
Repetition	789 (292)	97.0 (3.0)	989 (253)	96.7 (3.7)
Switch	981 (347)	95.2 (3.5)	1223 (349)	94.6 (3.3)
Mixing Cost	328 (230)	1.5 (2.5)	471 (220)	2.2 (3.6)
Switching Cost	192 (147)	1.8 (2.6)	234 (172)	2.1 (3.6)

Note: Standard deviations are presented in parentheses.

2.3.4.2 Mixing costs. Mixing costs represent the difference in performance on repetition trials in the mixed-task block and trials in the single-task blocks. The two groups did not differ in accuracy on either the repetition (*p* = .940) or single-task trials (*p* = .883). Additionally, there was no difference between the groups on the accuracy mixing cost (*p* = .960), though both groups showed a reliable cost (*p*s ≤ .041; Table 4). A three-way ANOVA on RTs with CTI length and trial type (repetition, single-task) as within-subjects factors and group as a between-subjects factor showed main effects for all three factors. Responses were faster on long compared to short CTI trials ($F(1,40) = 272.2$, $p < .001$, $\eta_p^2 = .872$) and on single-task trials compared to repetition trials ($F(1,40) = 132.1$, $p < .001$, $\eta_p^2 = .768$). These two factors interacted, showing a smaller mixing cost on trials with a long CTI than with a short CTI ($F(1,40) = 172.7$, $p < .001$, $\eta_p^2 = .812$). The interpreters were overall faster than the multilinguals ($F(1,40) = 6.5$, $p = .014$, $\eta_p^2 = .141$). Additionally, group interacted with trial type ($F(1,40) = 4.2$, $p = .046$, $\eta_p^2 = .096$; Figure 2), indicating a smaller mixing cost for interpreters. To verify that the difference in mixing cost was not due to the difference in overall speed we computed a proportional mixing cost ([repetition – single-task] / [repetition + single-task]). The interpreter advantage was maintained when considering this proportional mixing cost ($t(40) = 2.0$, $p = .045$, $d = .640$). The group factor also interacted with CTI length (F(1,40) = 4.2, p = .047, η_p^2 = .095), showing that the difference between the groups was larger with a short CTI (144 ms vs. 111 ms). The three-way interaction was not significant (p = .270).

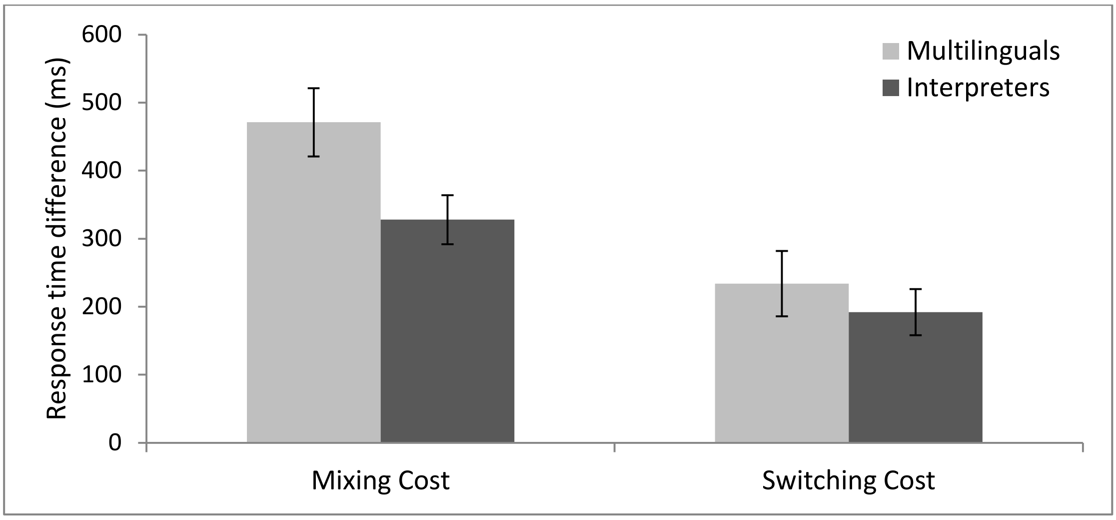

Figure 2: Mixing and switching costs by group. Error bars represent standard errors of the mean.

2.3.4.3 Regression analysis. To investigate the separate effects of memory and interpretation experience on the mixing cost, we completed a stepwise multiple regression analysis. Proportional mixing cost was selected as the dependent variable to mitigate differences in overall speed. Performance scores from the four tests of memory, group, and Raven's APM score were entered as potential variables (age did not warrant inclusion because its correlation with proportional mixing cost was not significant). A model including only spatial working memory explained 17.0% of the variance in mixing cost (F(1,40) = 9.3, p = .004). A second model, which added group as a predictor, explained 24.9% of the variance, significantly more than the first model (R^2 change = .095, F(1,39) = 5.1, p = .028; Table 5). No other variables significantly improved the model's predictive power. The switching cost was not correlated with any of the biographical or memory variables, thus we did not complete a regression analysis on these data.

Table 5: Multiple regression results

Variable	B	SE B	Β	*p*
Spatial WM	-.005	.002	-.435	.003
Group	.058	.025	.308	.028

2.4 Discussion

This study examined the memory and cognitive control abilities of professional simultaneous interpreters compared to a well-matched group of multilinguals. Memory was investigated using tests of short-term and working memory in the verbal and spatial domains. On these tests, the interpreters showed larger verbal STM, spatial STM, and verbal WM spans than the multilinguals. The color-word Stroop, ANT, and task-switching paradigms were employed to explore differences in cognitive control between the two groups. No group differences were seen in conflict resolution measured with the Stroop and ANT tasks. However, on the task-switching paradigm the interpreter group showed greater sustained control (i.e., smaller mixing cost) than the multilingual group, though no difference was seen in transient control (i.e., switching cost). A regression analysis on the mixing cost revealed that both spatial working memory and interpretation experience contributed to the magnitude of the mixing cost.

2.4.1 Are Interpreters Expert Bilinguals?

Previous studies have shown a bilingual advantage in transient control, measured as the switching cost, on a non-linguistic task-switching paradigm (Garbin et al., 2010; Prior & Gollan, 2011; Prior & Macwhinney, 2010; but see Hernández et al., 2013; Paap & Greenberg, 2013). This advantage is thought to stem from the fact that switching between tasks and switching between languages rely on the same domain-general processes. Bilinguals, then, have additional practice with these processes due to their experience switching between languages. The connection between language switching and task switching is reinforced by the results of Prior and Gollan (2011). That study demonstrated that bilinguals who engage more often in language switching show a benefit in transient control over bilinguals who switch less often (when corrected for differences in socioeconomic status). Interpreters, whose profession requires the use of multiple languages under demanding circumstances, could also be expected to show a benefit in transient control over non-interpreter multilinguals. The present study, however, demonstrated that while both professional interpreters and multilinguals show reliable switching costs, there is no difference between the two groups.

Bilingual advantages in conflict resolution have also been attributed to the need to manage multiple language sets, specifically the need to select between competing items. This advantage has typically been seen on Stroop tasks, the Simon task, and flanker paradigms (e.g., Bialystok, Martin, & Viswanathan, 2005; Costa et al., 2008; Hernández et al., 2010). As with transient control, a benefit in conflict resolution may be predicted for interpreters due to their management of languages in more demanding situations and greater need to produce target language output. For the most part, however, this prediction was not born out, as no group differences in the conflict effect were seen on the Stroop and ANT tasks in the current study. The interpreters, though, did show marginally faster overall response times, which have previously been linked to bilingual advantages (Bialystok, 2006; Costa et al., 2009).

Three previous studies examining conflict resolution in interpreters and multilinguals also found a general absence of differences between the groups. Yudes and colleagues (2011) examined monolinguals, bilinguals, and professional interpreters on the Simon task and found no differences between the groups (though that study also included a card sorting task which did show group differences). In a study examining the ANT, no differences in conflict resolution were seen between a group of professional intepreters and a group of bilinguals (though a interactive effect with the alerting and orienting networks was present; Morales, Padilla, Gómez-Ariza, & Bajo, 2015). In Köpke & Nespoulous (2006), professional interpreters, students of interpretation, and multilinguals were tested on color-word Stroop tasks in English and French. No differences between the groups were seen on the unilingual versions; though one of the bilingual versions did show an advantage for the students of interpretation. It should be noted that the paradigm used in that study did not allow for the calculation of the Stroop effect, where the bilingual advantage has generally been seen, but rather calculated the number of correct responses given in 45 seconds.

The lack of an enhanced benefit in interpreters could arise from two sources. First, while interpreters must control their languages under more demanding conditions, they may not actually spend more time switching between languages than other multilinguals. Indeed, comparisons of self-reported switching frequency revealed that in most situations the groups did not differ and when they did it was the the multilinguals who switched more often. It appears then that increased demands during switching may not result in enhanced practice. The second possibility is that interpreters have exercised their switching processes to a greater extent, but that enhancements are not linearly related to exercise or have a ceiling. In this case the extra practice would not add significantly to the enhancement that the interpreters already have due to their multilingual status. Regardless of the source, these results appear to indicate that interpreters do not continue to garner benefits in transient control and conflict resolution beyond that found in multilinguals. Thus, the data suggest that interpreters are not “expert bilinguals.”

2.4.2 Are Interpreters Unique Bilinguals?

The professional interpreters did distinguish themselves from the multilinguals in areas where bilingual benefits have not typically been seen. In particular, the interpreters displayed enhanced sustained control during the mixed-task block of the task-switching paradigm. This task block resembles simultaneous interpretation in that both require the maintenance of two (or more) “task sets.” In the case of the task-switching paradigm, these sets are the stimulus-response rules, whereas in simultaneous interpretation (SI) they are the input and output languages. Thus the interpreters’ enhancement may be due to the extensive practice they have in maintaining two task sets. This would further imply that interpreters recruit, at least in part, domain-general processes to keep both languages active. This account of the results, which appears to be supported by the inclusion of the group factor in the regression analysis, suggests that there is a direct effect of SI experience on the size of mixing costs. Alternatively, the difference in mixing cost between the groups could be due to the difference in their memory spans. Examination of the regression analysis revealed that spatial working memory and group membership were the only significant predictors of mixing cost. Interestingly, spatial WM is the only memory measure in which no interpreter advantage was seen. Thus, it may be that all four memory types influence the mixing cost, but that the other three are represented cumulatively with the group variable. This account would suggest that SI experience has an indirect effect on the size of mixing costs. Regardless of the path that the influence takes, SI experience is uniquely associated with enhanced sustained control.

2.4.3 Memory Effects

The interpreters additionally showed a specific advantage in verbal short-term and working memory as well as spatial short-term memory. In the verbal domain, the superiority of the interpreters replicates the findings of previous studies (e.g., Bajo et al., 2000; Christoffels et al., 2006; Köpke & Nespoulous, 2006; P. Padilla et al., 1995). Verbal memory is a critical component in simultaneous interpretation. Both short-term and working memory are burdened during SI to store content between input and output and rehearse pre-output translations. Further, as mentioned in the Introduction (section 2.1), SI requires simultaneous comprehension and production, and therefore may be considered a dual-task situation, similar to the working memory paradigm used in this study. Thus, the interpreter advantage in verbal memory is reasonable given the nature and demands of simultaneous interpretation.

The difference in spatial short-term memory, however, is the first evidence that interpreters have improved memory beyond the verbal domain as no previous study has included measures of non-verbal memory. At first blush this advantage in the spatial domain is surprising given the verbal nature of SI. The advantage could arise from increased general memory ability, analogous to the benefits seen in bilinguals on non-linguistic tasks of executive function (e.g., Costa et al., 2009, 2008; Garbin et al., 2010; Luk et al., 2011; Prior & Gollan, 2011; Prior & Macwhinney, 2010). However, the lack of an advantage on spatial working memory draws this explanation into question. Alternatively, the advantage may be rooted in the strategies used during SI. A distinction has been drawn between two interpreting strategies: a transcoding (or word-based) strategy and a meaning-based strategy (e.g., Anderson, 1994; Fabbro & Gran, 1994). In transcoding interpretation, the interpreter recodes individual words or multi-word units of the input into the target language. Recoding can occur at the phonological, morphological, syntactic, and semantic levels (Paradis, 1994). In all cases, though, verbal short-term memory would be employed to retain the content between input and output. In meaning-based interpretation, the interpreter fully comprehends the input, retains it at a non-verbal conceptual level, and then recodes the meaning in the target language for production (Fabbro & Gran, 1994). When using this strategy the short-term memory store which is taxed may be non-verbal. Thus, the benefit seen in spatial short-term memory could be directly related to the use of this meaning-based strategy. It is important to note that the two strategies are not mutually exclusive, so an individual interpreter may use both strategies and show associated benefits in both verbal and non-verbal short-term memory.

While it is clear that interpretation is associated with increased memory, it should be noted that the direction of causation remains unclear. Practice with interpretation may lead to increased memory, or individuals with better memory may self-select into the profession or be more successful in the long-run. Longitudinal studies offer the best possibility of clarifying the direction of effects. We are aware of one such study; however results from that study do not offer conclusive evidence. Macnamara and Conway (2013) examined students of ASL interpretation during their first and fourth semesters of training. Performance improved upon retesting on number-letter sequencing and backward digit span tasks, but no difference was seen on reading and operation span tasks. Future longitudinal studies may be of assistance in clarifying this point (see Chapter 3).

2.4.4 Working Memory and Mixing Costs

Finally, the influence of spatial working memory span on the mixing cost in the task-switching paradigm is in and of itself notable. Previous authors have theorized that mixing costs arise in part due to the greater memory load of the mixed-task blocks (Meiran, 1996; Rogers & Monsell, 1995). When directly explored by Rubin and Meiran (2005), however, this influence of memory load on the mixing cost was not confirmed. The present results approach this relation between working memory and mixing cost from the opposite side. That is, we show that individuals with larger working memory spans exhibit smaller mixing costs. These data support the notion that increased memory load contributes to mixing costs.

2.5 Concluding Remarks

The present study demonstrates that professional interpreters do not show bilingual benefits quantitatively beyond those seen in multilinguals. Instead, interpreters have a unique set of benefits that are related to the processes recruited during simultaneous interpretation. These benefits include increased verbal and spatial memory and enhanced sustained control. This specificity of the interpreter advantage to processes required during simultaneous interpretation echoes the finding of a recent study on the attentional control networks (Morales et al., 2015). Future studies in this area will greatly add to our knowledge, not only of the cognitive effects of simultaneous interpretation, but also of the cognitive effects of skill learning in adulthood generally.

CHAPTER 3

ARE INTERPRETERS BORN OR CREATED?

3.1 Introduction

The previous chapter demonstrated that interpreters show specific advantages in cognitive control compared to matched multilinguals. These advantages were seen in verbal short-term and working memory, as well as spatial short-term memory, and in sustained control in a task-switching paradigm. These findings add to a growing body of literature that has found interpreter advantages in memory (e.g., Bajo et al., 2000; Christoffels et al., 2006; Köpke & Nespoulous, 2006; P. Padilla et al., 1995) and articulatory suppression (e.g., Bajo et al, 2000, Kopke & Nespoulous, 2006; Padilla, Bajo, Macizo, 2012; Padilla et al., 1995; Yudes, Macizo, & Bajo, 2012), and more recently in some areas of executive control (Morales et al., 2015). Taken together these studies support the narrative that interpreters are not generally advantaged, but do exhibit enhancements in the areas of cognition recruited during simultaneous interpretation.

The provenance of these advantages, however, remains unclear. It may be that these abilities are acquired through training and/or experience with simultaneous interpretation. This understanding is supported by studies which show that targeted training in working memory and inhibitory control leads to improvements specific to the process trained (Berkman, Kahn, & Merchant, 2014; Melby-Lervåg & Hulme, 2013). Given that simultaneous interpretation places high demands on memory and other executive control processes, it may represent targeted training which could lead to enhancements in the specific areas utilized. Alternatively, the interpreter advantages may be due to innate characteristics that enable success in the field. Simultaneous interpretation is a highly selective career path. Beyond proficiency in a minimum of two or three languages, individuals must complete and pass a series of qualification exams to enter both training programs and professional associations. Thus, the individuals who consider and ultimately succeed in becoming professional interpreters may be those who naturally have the required abilities.

3.1.1 Previous Evidence

To examine the influence of SI training and experience on cognitive control, several studies have compared groups with different levels of SI experience. Padilla and colleagues examined professional interpreters in comparison with two groups of interpretation students (before and after courses on simultaneous interpretation) and non-interpreter controls (P. Padilla et al., 1995). The interpreters showed better performance than the other three groups on two tests of verbal memory (digit span and reading span). Additionally, the interpreters were unhindered by articulatory suppression in a verbal recall task, while the remaining three groups all showed a decrement in performance. These results suggest that the memory and articulatory suppression advantages are not innate, but rather acquired through experience with simultaneous interpretation.

This role of SI experience is partially upheld by another study which compared professional interpreters, students of interpretation, and monolinguals (Yudes et al., 2012). On a verbal recall task with articulatory suppression conditions of varying complexity the professional interpreters showed a decrease in recall only for pseudo-words in the high complexity condition. The students, on the other hand, were hindered by the high complexity condition regardless of item type (words and pseudo-words) and the monolinguals showed decrements in word and pseudo-word recall in all articulatory suppression conditions (standard and high complexity). When compared on a reading span task, however, the professional interpreters and students of interpretation did not differ and both groups performed better than the monolinguals.

A third study examining professional interpreters, students of interpretation, bilingual controls, and monolingual controls further complicates the pattern of effects (Köpke & Nespoulous, 2006). In this study no differences were seen across the four groups on tasks of verbal short-term memory (digit span and word span). On a verbal working memory task (listening span), though, the students of interpretation outperformed the two control groups, while the professional interpreters did not differ from any of the groups. A similar pattern of effects was evidenced in a verbal recall task under articulatory suppression. Finally, this study also examined the color-work Stroop task using English, French, and bilingual versions. The students of interpretation provided more correct responses in 45 seconds than the professional interpreters and bilinguals on one of the bilingual versions (words written in L2 English, respond in L1 French); no differences were seen on the other three versions.

A final study assessed students of interpretation at the close of their first and second years, as well as bilingual controls, on tasks of verbal memory (Tzou, Eslami, Chen, & Vaid, 2011). On a short-term memory task (digit span), the second year students performed better than the bilingual controls, while the first year students did not differ from either group. On a working memory task (reading span), however, both student groups outperformed the bilingual control group.

These four studies present a mixed view of the influence of SI training and experience on cognitive control. Three of the studies provide some support for an effect of SI training on memory and articulatory suppression effects (P. Padilla et al., 1995; Tzou et al., 2011; Yudes et al., 2012). However, support is also available for innate differences (Tzou et al., 2011) and the influence of SI experience (Yudes et al., 2012). Finally, one of the studies failed to find an interpreter advantage (Köpke & Nespoulous, 2006). These studies must be interpreted with caution, though, as they all employed a cross-sectional design. In this situation, when comparing professionals to students, there is an inherent age difference which may be contributing to the effects. In particular, the advantages for students seen in the study by Köpke and Nespoulous could be explained by age-related changes in cognition.

Longitudinal studies of individuals learning to interpret offer a better method to explore the origin of the interpreter advantages. We are aware of one longitudinal study that investigates changes in cognitive control in students of interpretation. Macnamara and Conway (2013) examined students of ASL (American Sign Language) interpretation during their first and fourth semesters of training. These students evidenced improvements on a task-switching paradigm, the Wisconsin Card Sorting Test, number-letter sequencing, and backward digit span tasks; however, no change was seen on reading and operation span tasks. While the authors of this study did use a longitudinal design, they did not include a control group. Thus, it is not clear if the improvements seen were due to training in interpretation or rather repetition of the task (e.g., learning effects). Additionally, ASL interpretation may be qualitatively different than interpretation between two spoken languages. As there is only one spoken language stream in ASL interpretation, the level of interference between the languages is likely lower. Indeed, evidence from studies of bimodal bilinguals (one spoken language and one signed language) suggests that these individuals do not exhibit the same level of cognitive enhancements that unimodal bilinguals show (e.g., Emmorey, Luk, Pyers, & Bialystok, 2008).

3.1.2 The Present Study

To understand the changes in cognition associated with training in simultaneous interpretation we examined students earning a Master in Conference Interpreting at the beginning and end of their two-year program. Further, to address the confounds related to the passage of time and repetition of tasks, we compared this group of interest to two control populations. The first were students earning a Master in Translation. Similar to the students of interpretation, these students spoke multiple languages and were engaged in language improvement and high level language usage during the intervening two years. Critically, though, the students of translation did not learn the simultaneous interpretation skill. Thus, the comparison of these two groups specifically targets training in simultaneous interpretation. The second control group consisted of monolinguals studying non-language fields. The inclusion of this group allowed us to tease apart the potential effects of increased multilingualism from the general time and repetition effects.

All three groups were tested longitudinally on the same battery of tasks used in Chapter 2 to examine professional interpreters and multilinguals. Based on the results of that study, we expected to see differences between the groups in verbal short-term and working memory, spatial short-term memory, and sustained control (measured with the mixing cost in the task-switching paradigm). How these differences evolved in relation to the training period would speak to the question of innate versus trained abilities. A role for innate abilities would be supported by better performance in the students of interpretation compared to the other groups, across the training period. Alternatively, a role of training would be supported by an increase in performance in the students of interpretation, but not the two control groups. Finally, a lack of differences between the groups at either time point may suggest a role of SI experience beyond the two year training period. It should be noted that the abilities are not intrinsically linked, thus some abilities may be innate and others trained.

3.2 Methods

3.2.1 Participants

Three groups of participants were examined longitudinally in this study. The first group was composed of students earning a Master degree in Conference Interpreting at the University of Trieste (one student earned a Master in Conference Interpreting at another Italian university). These students were tested at the start of their Master's program (within the first three months of courses for all but one participant who was tested four months before the program started) and again approximately two years later at the conclusion of the program (mean time between phases = 22.7 months, SD = 1.4, range 19.5-24.4). Twenty-two students completed the first phase of which seventeen (12 females) returned for the second phase. The second group consisted of students earning a Master degree in Translation at the University of Trieste. As with the first group, these students were tested at the start of their program and approximately two years later at the conclusion of the program (mean time between phases = 24.1 months, SD = 1.0, range 22.9-25.7). Twenty-one students completed the first phase and ten of these students (8 females) returned for the second phase. The third group of participants were monolingual controls who were studying non-language subjects at the University of Trieste. These participants were tested twice with approximately two years between the phases (mean time between phases = 23.7 months, SD = 1.4, range 21.9-26.2). While the specific time of testing was not linked to their academic calendar, all participants were full-time students during the intervening two years. Nineteen monolinguals participated in the first phase and eleven (4 females) returned for the second phase. All participants were right-handed and had no known neurological or psychiatric problems. They all reported normal color vision, which was confirmed with the Ishihara Color Vision Test (Ishihara, 1972). Further, the three groups did not differ in age (p = .128) or socioeconomic status (measured as mother's years of education; p = .865), however, the monolingual group did have approximately one more year of education ($F(2,35) = 3.846$, p = .031; Table 6). Finally, while the study attrition rates for the students of translation and the monolingual controls were higher than for the students of interpretation, this difference does not reflect attrition from the participants' programs of study. Instead, at least for the students of translation, it was due to a difference in average thesis preparation time. This meant that the students of translation moved away from the Trieste area earlier than the students of interpretation, so fewer were available for re-testing at the specified time. Participants gave written informed consent and were compensated for their time. The study was approved by the ethical committee of the Scuola Internazionale Superiore di Studi Avanzati (SISSA).

3.2.2 Tasks and Procedure

Participants were tested individually in a sound-attenuated booth. During the first phase of testing participants completed a language history questionnaire, the Wechsler Adult Intelligence Scale (WAIS-R), four tests of memory, and four tests of executive functioning (two additional tasks were included but not analyzed here, a foreperiod task and the training for a novel inductive reasoning task, see Babcock & Vallesi, 2015). These tasks were split across two testing sessions. In the second phase, participants repeated the memory and executive functioning tasks and updated their language history information in a single session.

3.2.2.1 Language history questionnaire. Participants were asked to provide information about all of the languages they knew and/or studied. For each language they were asked to detail how and when they learned the language, including immersion experiences, as well as to provide a self-rating in the areas of reading, writing, speaking, and understanding on a 5-point Likert scale. The participants also completed a questionnaire developed to identify functional fluency in non-native languages. Functional fluency was operationalized as a B2 level or above in the Common European Framework of Reference for Languages (CEF). The questionnaire asked participants to give their CEF level and respond to eight yes-or-no questions which targeted the B1-B2 border (see Appendix A for questionnaire items). The questionnaire contained two items for each of the four abilities (reading, writing, speaking, and oral comprehension), one item focused on academic usage and the other on personal usage. Participants were considered functionally fluent in languages for which they responded yes to seven or eight items. This questionnaire was used to screen the Monolingual controls at Phase 1. Interested participants who were deemed functionally fluent in a language other than Italian were not invited to participate. Additionally, all participants completed this questionnaire at Phase 2.

To examine potential differences in language proficiency between the Interpretation and Translation students we compared their self-ratings at Phase 1 and Phase 2. However, given that participants were asked to report all languages they had studied, even for brief periods, an average across all non-native languages would not have been an accurate representation of their abilities. Instead, the self-ratings were averaged across the two or three languages each participant was studying as part of their Master's program. These comparisons identified an advantage for Interpretation students in speaking ability at Phase 1 ($t(25) = 2.756$, $p = .011$), but there were no other differences between the groups at either Phase 1 or Phase 2 ($ps \geq .222$; Table 6). Additionally, we compared the number of functional languages at Phase 2 between these groups and found no difference ($t(25) = 1.423$, $p = .167$).

Table 6: Biographical and language characteristics of the three participant groups

	Interpretation students (N=17)	Translation students (N=10)	Monolingual controls (N=11)
Age at Phase 1 (in years)	22.5 (0.6)	24.2 (3.6)	23.5 (1.9)
Years of education at Phase 1	16.2 (0.4)	16.5 (0.7)	17.4 (1.9)
Verbal IQ	123 (8)	113 (10)	114 (13)
Performance IQ	113 (15)	111 (9)	113 (11)
Mother's years of education	13.6 (4.0)	13.1 (3.2)	12.9 (2.6)
Number of functional languages at Phase 2	3.6 (0.7)	3.2 (0.6)	1.5 (0.5)
Phase 1 averaged reading level*	4.3 (0.4)	4.2 (0.4)	
Phase 1 averaged writing level*	3.7 (0.3)	3.5 (0.5)	
Phase 1 averaged speaking level*	3.7 (0.5)	3.2 (0.5)	
Phase 1 averaged understanding level*	4.1 (0.4)	4.1 (0.5)	
Phase 2 averaged reading level*	4.5 (0.5)	4.5 (0.5)	
Phase 2 averaged writing level*	3.9 (0.5)	4.0 (0.4)	
Phase 2 averaged speaking level*	4.0 (0.6)	3.8 (0.4)	
Phase 2 averaged understanding level*	4.4 (0.4)	4.3 (0.4)	

Note: Values in parentheses are standard deviations. *These values were averaged across the two or three languages each participant studied as part of their Master's program.

3.2.2.2 Wechsler Adult Intelligence Scale (WAIS-R). The WAIS-R (Wechsler, 1981) was selected to measure intelligence because it has an Italian translation which has been normed on an Italian population (Orsini & Laicardi, 1997).

The test is composed of eleven subtests, which are divided into tests of verbal intelligence and tests of performance intelligence. Thus, the WAIS-R allows for the separate calculation of verbal IQ and performance IQ. The verbal tests included Information, Comprehension, Arithmetic, Digit Span, Similarities, and Vocabulary. The performance tests included Picture Arrangement, Picture Completion, Block Design, Object Assembly, and Digit Symbol. Comparing the verbal and performance IQs across the three groups revealed no difference in performance IQ (p = .936), however the Interpretation students had a higher verbal IQ than the other two groups ($F(2,35) = 4.159$, p = .024).

3.2.2.3 Memory tests. Short-term and working memory were assessed in the verbal and spatial domains using the same tasks as employed in Chapter 2 (see Section 2.2.2.4). The letter span and matrix span tasks (Kane et al., 2004) were used to examine short-term memory in the verbal and spatial domains, respectively. The structure of these tasks was identical to that described in Chapter 2; however, the instructions and text throughout were translated into Italian. Additionally, a different set of twelve consonants was used in the letter span task. The original version included three consonants which are not used in the Italian language (J, K, and Y). To avoid confounds related to familiarity with these letters, they were replaced with Italian consonants (C, V, and Z). The selection of the replacements was completed so that the English letter frequency (Lewand, 2000) in the English version and the Italian letter frequency (Singh & Galli, 1999) in the Italian version were similar. To assess working memory in the verbal and spatial domains the automated operation span task and the automated symmetry span task were employed (Unsworth et al., 2005). These tasks were also translated versions of those used in Chapter 2 (Section 2.2.2.4).

Additionally, the same letter replacements used in the letter span task were completed in the operation span task.

3.2.2.4 Tests of executive functioning. The four tasks of executive functioning examined in this study were the color-word Stroop task, a spatial Stroop task, the Attention Network Test (ANT), and a non-linguistic task-switching paradigm.

3.2.2.4.1 Color-word Stroop task. This task was a shortened version of the task presented in Puccioni and Vallesi (2012b). The task parameters were identical to those used in Chapter 2 (see Section 2.2.2.1). The only differences were the language of the task, which was Italian in the present study, and the response pad used, a Cedrus RB-730 in the present study.

3.2.2.4.2 Spatial Stroop task. This task was a shortened version of the task presented in Puccioni and Vallesi (2012a). Stimuli consisted of arrows pointing to the upper-left, upper-right, lower-left, and lower-right presented in one of the four quadrants of the screen. Participants were asked to ignore the position and indicate the direction of the arrow through a key press. Responses were made on a keyboard using the following keys: R for upper-left, O for upper-right, V for lower-left, and M for lower-right. Participants were asked to respond as quickly and accurately as possible. As in the color-word Stroop task, stimuli were categorized as congruent (e.g., upper-right pointing arrow in the upper-right quadrant) or incongruent (e.g., upper-right pointing arrow in the lower-right quadrant). The details of the task and training procedures were identical to the ones used in the color-word Stroop task.

3.2.2.4.3 ANT. This task was a translated version of the ANT task used in Chapter 2 (see Section 2.2.2.2), which was adapted from Costa and colleagues (2008). Additionally, a Cedrus RB-730 response pad was used.

3.2.2.4.4 Task-switching paradigm. This task was a translated version of the task-switching paradigm used in Chapter 2 (see Section 2.2.2.3), which was modified from Rubin and Meiran (2005). As in the previous tasks, a Cedrus RB-730 response pad was used.

3.3 Results

Below we present the results from the four tests of memory and four executive functioning tasks. Given the small number of participants in each of the three groups, we began all analyses by comparing the Interpretation students and the Translation students. We chose this comparison because it is the most informative about the effects of training in simultaneous interpretation, while controlling for potential confounds of high proficiency in multiple languages. In cases where changes over time were seen, but did not differ between these two groups we then combined these groups and compared them to the monolingual control group. These analyses allowed us to tease apart the effects of task repetition from improvements related specifically to multilingualism.

3.3.1 Memory Tests

Mixed-effects ANOVAs with phase (first, second) as a within-subjects factor and group (Interpretation students, Translation students) as a between-subjects factor were computed on the score of each of the four memory tasks and the number of errors on the two working memory tasks.

The letter span task revealed a main effect of phase ($F(1,25) = 7.837$, $p = .010$, $\eta_p^2 = .239$), as well as an interaction between phase and group ($F(1,25) = 4.533$, $p = .043$, $\eta_p^2 = .153$; Figure 3). Post hoc *t*-tests (evaluated at $\alpha = .025$ to correct for multiple comparisons) revealed a significant increase in letter span score at Phase 2 in the Interpretation students ($t(16) = 3.655$, $p = .002$), but no difference between the phases in the Translation students ($p = .597$). Further, the two groups showed no difference in letter span at Phase 1 ($p = .527$), but a significant difference at Phase 2 ($t(25) = 2.501$, $p = .019$).

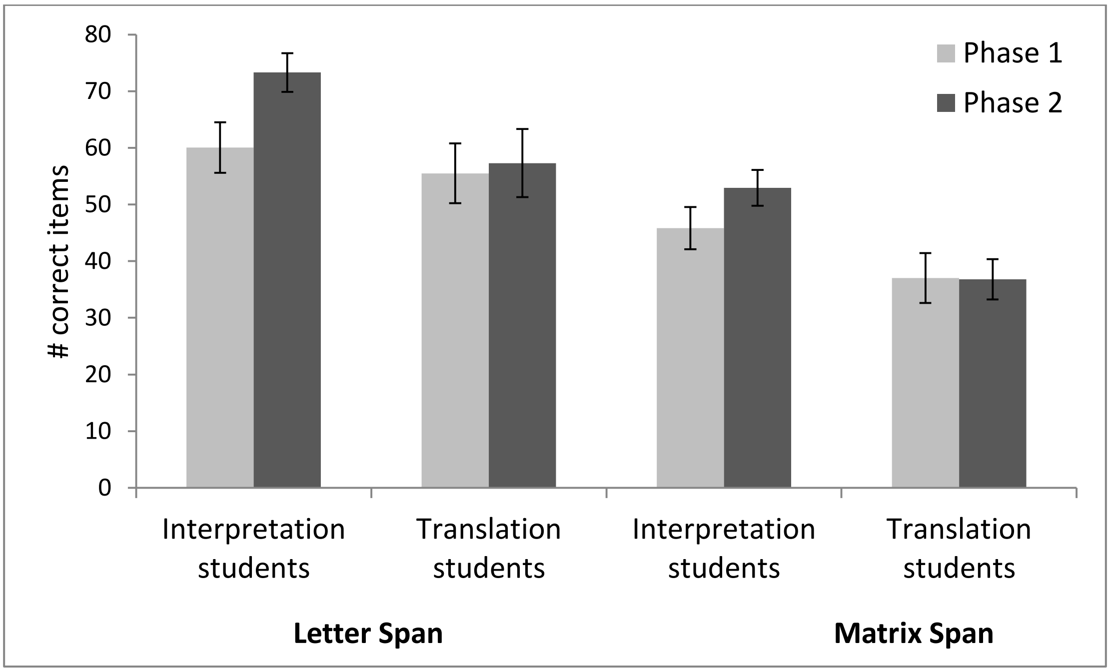

Figure 3: Performance on tests of short-term memory by group and phase. Error bars represent standard errors of the mean. The number of items, and therefore maximum score, for the memory tests were as follows: Letter Span – 99; Matrix Span – 81.

The matrix span task revealed a main effect of group (F(1,25) = 6.383, p = .018, η_p^2 = .203) and trends toward a main effect of phase and an interaction between phase and group (F(1,25) = 2.241, p = .147, η_p^2 = .082 and F(1,25) = 2.508, p = .126, η_p^2 = .091, respectively; Figure 3). A decomposition of this potential interaction reveals the same pattern as in the letter span task. The Interpretation students significantly increased their matrix span score (t(16) = 3.176, p = .006), while the Translations students did not (p = .967). Additionally, a difference between the groups was seen at Phase 2 (t(25) = 3.275, p = .003), but not at Phase 1 (p = .149).

Analyses on the score on the operation span task revealed a main effect of phase (F(1,25) = 7.278, p = .012, η_p^2 = .225) and a trending interaction between phase and group (F(1,25) = 2.057, p = .164, η_p^2 = .076; Figure 4). The Interpretation students showed an increase in operation span between the phases (t(16) = 3.724, p = .002), while the Translation students did not (p = .503). However the two groups did not differ at either time point (ps $\geq$ .483). Analyses on the number of errors in the operation span task revealed no significant main effects (ps $\geq$ .662), but there was an interaction of phase and group (F(1,25) = 6.702, p = .016, η_p^2 = .211). Splitting the errors into incorrect responses on the intervening arithmetic operation and responses to the operation beyond the time limit determined by each participant's performance in the training revealed that the effect was due to differences in the number of incorrect responses on the intervening operation task (F(1,25) = 5.543, p = .027; Figure 4). The Interpretation students showed a significant decrease in errors (t(16) = 2.710, p = .015), while the Translation students showed a non-significant increase (p = .369).

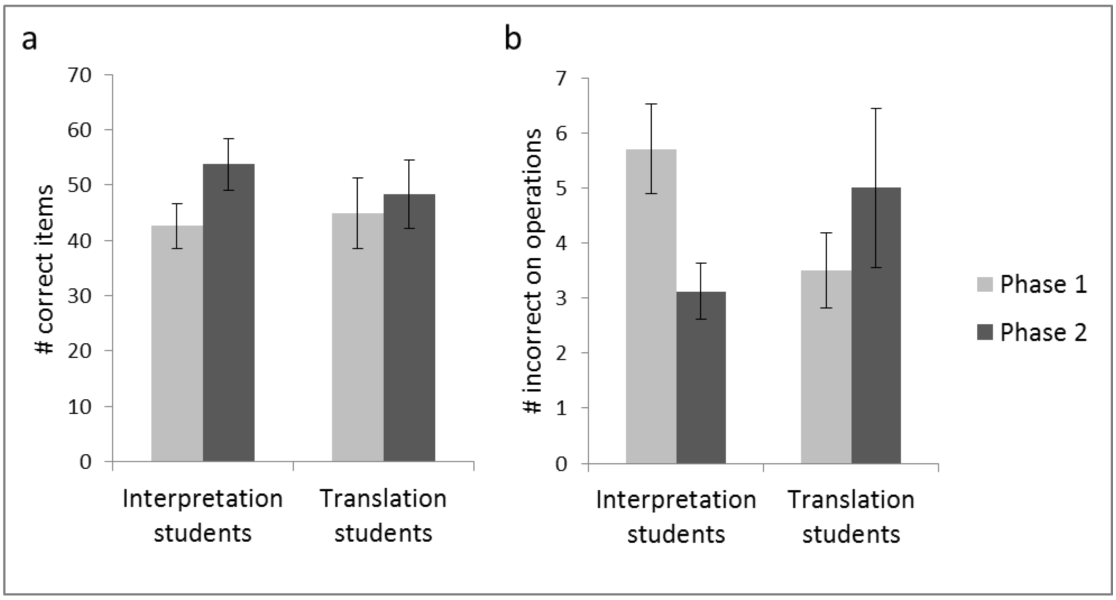

Figure 4: (a) items recalled and (b) errors on the intervening task on the operation span task by group and phase. Error bars represent standard errors of the mean. The number of items, and therefore maximum score, was 75.

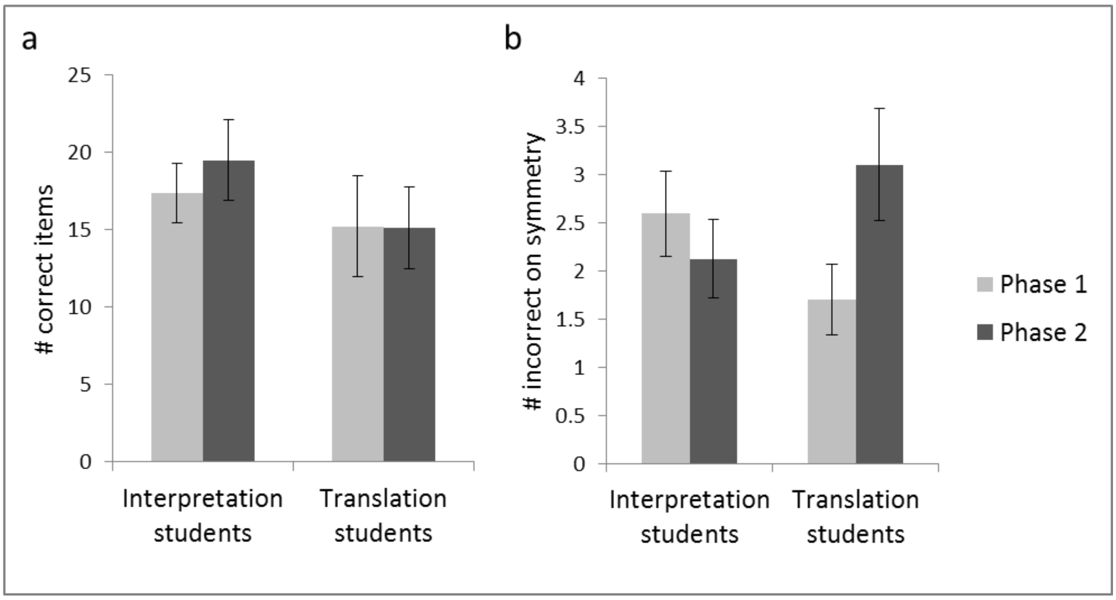

Figure 5: (a) items recalled and (b) errors on the intervening task on the symmetry span task by group and phase. Error bars represent standard errors of the mean. The number of items, and therefore maximum score, was 42.

Analyses on the score on the symmetry span task revealed no significant main effects or interaction (*p*s ≥ .303; Figure 5). Analyses on the number of errors also revealed no main effects or interaction (*p*s ≥ .517). Given the results concerning the number of incorrect responses to the intervening task in the operation span task, we conducted the same analysis on the symmetry span data. As in the operation span task, there was an interaction of phase and group ($F(1,25) = 4.419$, $p = .046$, $\eta_p^2 = .150$; Figure 5). This effect was due to a non-significant decrease among the Interpretation students ($t(16) = .836$, $p = .416$) and a trending increase among the Translation students ($t(9) = 2.143$, $p = .061$).

As the effects seen in the number of incorrect responses on the operation span and symmetry span tasks were due to increases in Translation students, in addition to decreases in Interpretation students, we decided to examine these effects in the Monolingual controls as well. With these extra analyses we can understand whether learning to translate had a specific effect on the number of incorrect responses to the intervening task. Two mixed effects ANOVAs with phase (first, second) as a within-subjects factor and group (Interpretation students, Monolingual controls) as a between-subjects factor were computed on the number of incorrect responses to the intervening task in the operation span and symmetry span tasks. The analysis on the operation span task yielded a significant interaction between phase and group ($F(1,26) = 11.209$, $p = .002$, $\eta_p^2 = .301$), with the Monolingual controls showing a significant increase in incorrect responses on the intervening task ($t(10) = 2.375$, $p = .039$). The symmetry span analysis, however, showed no significant interaction ($p = .245$), though numerically the Monolingual controls increased their number of incorrect responses on the symmetry judgments.

3.3.2 Color-Word Stroop Task

One participant (a male Interpretation student) was identified as an extreme outlier based on his accuracy rate during Phase 1 (more than 3 interquartile ranges below the 1st quartile) and therefore excluded from all analyses. The first trial in each block was not considered for all analyses on accuracy. Non-parametric tests were used to analyze the accuracy data given that these data were not normally distributed. The Mann-Whitney U test was used to compare the two groups and the Wilcoxon signed-ranked test was used to compare the within-subjects factors. Response time (RT) data were analyzed using a mixed effects three-way ANOVA with phase (first, second) and trial type (congruent, incongruent) as within-subjects factors and group (Interpretation students, Translation students) as a between-subjects factor. For these data the first trial in each block and error trials were excluded. Additionally, for each participant, trials with an RT more than 3 standard deviations (SD) from their individual mean were excluded. Finally, trials following an error were excluded to avoid post-error slowing confounds (Burns, 1965).

Analyses on the accuracy data revealed that the groups did not differ in accuracy at either Phase 1 (p = .660) or Phase 2 (p = .262). Additionally, there were no differences between the groups on congruent or incongruent trials in either phase (ps ≥ .150). Across the groups there was a significant decrease in accuracy from Phase 1 to Phase 2 (Wilcoxon T = 56.5, Z = 2.682, p = .007). The analysis on the RT data revealed a main effect of phase ($F(1,24) = 20.070$, $p < .001$, $\eta_p^2 = .455$; Table 7), with faster responses at Phase 2 than at Phase 1. Additionally, there was a main effect of trial type ($F(1,24) = 40.834$, $p < .001$, $\eta_p^2 = .630$) due to faster responses to congruent trials. No other effects or interactions were significant (ps ≥ .187).

Table 7: Color-word Stroop: mean response times and accuracy by group and phase

	Interpretation students		Translation students		Monolingual controls	
	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)
Phase 1						
Congruent	609 (93)	98.2 (1.4)	613 (69)	96.9 (4.5)	653 (148)	97.5 (2.8)
Incongruent	654 (114)	96.8 (3.7)	698 (133)	96.5 (2.8)	724 (147)	96.6 (4.3)
Stroop effect	45 (47)	1.4 (3.8)	85 (84)	0.4 (3.2)	71 (69)	0.9 (4.9)
Phase 2						
Congruent	561 (86)	96.9 (3.2)	585 (81)	95.0 (6.6)	581 (89)	96.5 (2.6)
Incongruent	617 (84)	94.9 (4.9)	656 (117)	92.4 (5.6)	658 (109)	93.9 (4.4)
Stroop effect	56 (30)	1.9 (4.5)	70 (62)	2.6 (3.4)	77 (55)	2.6 (4.8)

Note: Values in parentheses are standard deviations.

To determine if the faster response times at Phase 2 as compared to Phase 1 were due to language factors or repetition of the task, the same analyses were completed comparing the Interpretation and Translation students together to the Monolingual controls. The decrease in both accuracy (Wilcoxon T = 83.5, Z = 3.539, $p < .001$) and response time ($F(1,35) = 30.870$, $p < .001$, $\eta_p^2 = .469$) between Phase 1 and Phase 2 was repeated in this analysis. Additionally, there was a trending interaction between phase and group ($F(1,35) = 2.286$, $p = .139$, $\eta_p^2 = .061$), due to a larger difference between the phases in the Monolingual controls. No other effects or interactions were significant (ps $\geq .361$).

3.3.3 Spatial Stroop Task

Two participants (one female Interpretation student and one female Translation student) were identified as extreme outliers based on their accuracy rates during Phase 1 (more than 3 interquartile ranges below the 1st quartile) and therefore excluded from all analyses. The first trial in each block was not considered for all analyses on accuracy. Non-parametric tests were used to analyze the accuracy data given that these data were not normally distributed. The Mann-Whitney U test was used to compare the two groups and the Wilcoxon signed-ranked test was used to compare the within-subjects factors. Response time data were analyzed using a mixed effects three-way ANOVA with phase (first, second) and trial type (congruent, incongruent) as within-subjects factors and group (Interpretation students, Translation students) as a between-subjects factor. For these data the first trial in each block and error trials were excluded. Additionally, for each participant, trials with an RT more than 3 standard deviations (SD) from their individual mean were excluded. Finally, trials following an error were excluded to avoid post-error slowing confounds (Burns, 1965).

Analyses on the accuracy data revealed that the groups did not differ in accuracy at either Phase 1 (p = .846) or Phase 2 (p = .803). Additionally, there were no differences between the groups on congruent or incongruent trials in either phase (ps ≥ .301). Finally, across the groups there was no difference in accuracy between Phase 1 and Phase 2 (p = .693).

The analysis on the RT data revealed main effects of phase ($F(1,23) = 21.383$, $p < .001$, $\eta_p^2 = .482$) and trial type ($F(1,23) = 175.222$, $p < .001$, $\eta_p^2 = .884$). These effects were due to faster responses in Phase 2 and to congruent trials. Additionally, these factors interacted ($F(1,23) = 6.876$, $p = .015$, $\eta_p^2 = .230$), indicating a larger decrease in RT on incongruent than congruent trials across the phases. There was no main effect of group (p = .978) and this factor did not interact with phase (p = .851). However, the trial type by group interaction was trending ($F(1,23) = 2.268$, $p = .146$, $\eta_p^2 = .090$), as was the phase by trial type by group interaction ($F(1,23) = 2.917$, $p = .101$, $\eta_p^2 = .113$, Table 8). Post hoc t-tests (evaluated at $\alpha = .025$ to correct for multiple comparisons) revealed a significant decrease in the Stroop effect (the difference between congruent and incongruent trials) at Phase 2 in the Interpretation students (t(15) = 3.138, p = .007), but no difference between the phases in the Translation students (p = .303). However, the groups did not differ in Stroop effect size at Phase 2 (p = .494), and the Interpretation group showed a numerically larger Stroop effect at Phase 1 (t(23) = 1.779, p = .089).

Table 8: Spatial Stroop: mean response times and accuracy by group and phase

	Interpretation students		Translation students		Monolingual controls	
	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)
Phase 1						
Congruent	514 (99)	99.3 (1.0)	525 (83)	99.1 (2.1)	523 (102)	99.6 (0.7)
Incongruent	622 (134)	94.2 (5.4)	603 (106)	95.4 (3.2)	639 (164)	93.9 (6.7)
Stroop effect	108 (46)	5.1 (5.1)	78 (32)	3.8 (2.8)	115 (86)	5.6 (6.3)
Phase 2						
Congruent	464 (66)	98.6 (1.3)	469 (50)	98.7 (2.3)	452 (81)	99.3 (1.1)
Incongruent	542 (79)	95.4 (3.4)	540 (54)	95.7 (4.5)	547 (124)	91.1 (6.6)
Stroop effect	79 (26)	3.2 (3.5)	71 (25)	3.1 (5.7)	94 (71)	8.1 (6.1)

Note: Values in parentheses are standard deviations.

To determine if the faster responses and smaller Stroop effect at Phase 2 as compared to Phase 1 were due to language factors or repetition of the task, the same analyses were completed comparing the Interpretation and Translation students together to the Monolingual controls. The results of the RT analysis replicated the previous findings, showing main effects of phase (F(1,34) = 41.257, p < .001, η_p^2 = .548) and trial type (F(1,34) = 114.818, p < .001, η_p^2 = .772), as well as an interaction between these two factors (F(1,34) = 11.343, p = .002, η_p^2 = .250). However, there was no main effect of group, nor any significant interactions with the group factor (ps ≥ .315).

3.3.4 ANT

Two participants (one male and one female Translation student) were identified as extreme outliers based on their accuracy rate (more than 3 interquartile ranges below the 1st quartile) and therefore excluded from all analyses. For all analyses on accuracy, the first trial in each block was not considered. Since the accuracy data were non-normally distributed, non-parametric tests were used for their analysis. Wilcoxon Signed Rank tests were used to compare conditions within each group and Mann-Whitney U tests were used to compare conditions between the two groups. For the analysis of response time data the first trial in each block and error trials were excluded. Additionally, for each participant, trials with an RT more than 3 SDs from their individual mean were excluded. Finally, trials following an error were excluded to avoid post-error slowing confounds (Burns, 1965).

3.3.4.1 Conflict effect. The conflict effect analyzes the difference between congruent and incongruent trials. To examine this effect, RT data were analyzed using a mixed effects three-way ANOVA with phase (first, second) and trial type (congruent, incongruent) as within-subjects factors and group (Interpretation students, Translation students) as a between-subjects factor.

Analyses on the accuracy data revealed that the groups did not differ in accuracy at Phase 1 (p = .549), however the Translation students had higher accuracy at Phase 2 (Mann-Whitney U = 101.5, z = 1.966, p = .049). Further, the Interpretation students showed a significant decrease in accuracy from Phase 1 to Phase 2 (Wilcoxon T = 12, Z = 2.906, p = .004), while the Translation students showed no change (p = .608). The analysis on the RT data revealed a main effect of phase (F(1,23) = 16.071, p = .001, η_p^2 = .411; Table 9), with faster responses at Phase 2 than at Phase 1. They also revealed a main effect of trial type (F(1,23) = 360.814, p < .001, η_p^2 = .940) due to faster responses to congruent trials. No other effects or interactions were significant (ps ≥ .209).

To determine if the faster response times at Phase 2 were due to language factors or repetition of the task, the same analyses were completed comparing the Interpretation and Translation students together to the Monolingual controls. As with the previous RT analysis, there were main effects of phase ($F(1,34) = 21.454$, $p < .001$, $\eta_p^2 = .387$) and trial type ($F(1,34) = 616.265$, $p < .001$, $\eta_p^2 = .948$). Additionally, there was an interaction between trial type and group ($F(1,34) = 6.738$, $p = .014$, $\eta_p^2 = .165$), due to a larger conflict effect for the Monolingual controls. All other effects were non-significant (*p*s $\geq .478$).

3.3.4.2 Alerting effect. The alerting effect was defined as the difference between trials cued with the double cue and those with no cue. This effect was analyzed using a mixed effects two-way ANOVA with phase (first, second) as a within-subjects factor and group (Interpretation students, Translation students) as a between-subjects factor. This analysis revealed no main effects of phase or group ($p = .872$ and $p = .860$, respectively), however there was a significant interaction between phase and group ($F(1,23) = 6.866$, $p = .015$, $\eta_p^2 = .230$). This interaction was due to a significant increase in the alerting effect among the Interpretation students ($t(16) = 2.522$, $p = .023$) and a non-significant decrease among the Translation students ($p = .200$).

3.3.4.3 Orienting effect. The orienting effect was defined as the difference between central cue trials and spatial cue trials. This effect was examined using a mixed effects two-way ANOVA with phase (first, second) as a within-subjects factor and group (Interpretation students, Translation students) as a between-subjects factor. The analysis revealed no significant effects (*p*s $\geq .255$) though a reliable difference between spatially and centrally cued trials was seen across the groups and phases ($F(1,24) = 9.742$, $p = .005$, $\eta_p^2 = .298$).

Table 9: ANT: mean response times and accuracy by group and phase

	Interpretation students		Translation students		Monolingual controls	
	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)
Phase 1						
Congruent	444 (85)	99.0 (1.1)	485 (86)	99.3 (0.7)	428 (42)	98.9 (1.3)
Incongruent	507 (86)	96.8 (2.7)	513 (82)	96.7 (4.9)	503 (42)	94.0 (5.4)
Conflict effect	63 (19)	2.2 (2.6)	65 (20)	2.6 (4.7)	75 (18)	4.8 (5.2)
Alerting effect	6 (13)	0.9 (1.6)	16 (20)	0.2 (1.8)	5 (14)	0.4 (4.4)
Orienting effect	12 (18)	1.6 (2.3)	1 (26)	0.2 (2.3)	13 (11)	-0.1 (2.4)
Phase 2						
Congruent	387 (39)	98.6 (1.4)	394 (45)	99.0 (1.0)	385 (49)	98.2 (1.7)
Incongruent	445 (43)	92.1 (7.0)	456 (46)	97.1 (4.4)	457 (46)	88.6 (10.9)
Conflict effect	58 (16)	6.5 (6.3)	62 (17)	1.9 (3.8)	72 (21)	9.5 (10.3)
Alerting effect	16 (16)	-0.5 (2.7)	8 (11)	-0.2 (2.6)	2 (14)	0.1 (3.3)
Orienting effect	9 (13)	0.4 (3.1)	7 (14)	-0.8 (1.7)	5 (12)	0.9 (4.0)

Note: Values in parentheses are standard deviations.

3.3.5 Task-Switching Paradigm

For all analyses on accuracy the first trial in each block and the first trial after the rest break in block 3 were not considered. Since the accuracy data were non-normally distributed, non-parametric tests were used for their analysis. Wilcoxon Signed Rank tests were used to compare conditions across the groups and Mann-Whitney U tests were used to compare conditions between the two groups. For the analysis of response time data the first trial in each block, the first trial after the rest break in block 3, and error trials were excluded. Additionally, for each participant, trials with an RT more than 3 SDs from their block-type mean (single-task, mixed-task) were excluded. Finally, trials following an error were excluded to avoid post-error slowing confounds (Burns, 1965) and since these trials are likely to be mis-categorized as switch or repetition trials.

3.3.5.1 Switching costs. Switching costs represent the difference in performance on switch trials and repetition trials in the mixed-task block. The

Interpretation students showed marginally more accurate responses to switch trials at Phase 2 than the Translation students (Mann-Whitney U = 49.5, z = 1.785, p = .074). No other group differences in accuracy were significant (ps ≥ .414). Across the groups, accuracy was higher to repetition trials than to switch trials at both Phase 1 and Phase 2 (ps ≤ .001). A three-way ANOVA on RTs with phase (first, second) and trial type (repetition, switch) as within-subjects factors and group (Interpretation students, Translation students) as a between-subjects factor was computed. Main effects of phase (F(1,25) = 39.842, p < .001, η_p^2 = .614) and trial type (F(1,25) = 60.541, p < .001, η_p^2 = .708) were evident due to faster responses at Phase 2 and to repetition trials. Additionally, these factors showed a marginal interaction (F(1,25) = 3.094, p = .091, η_p^2 = .110; Table 10), revealing that switching costs also decreased between Phase 1 and Phase 2. No effects with group were significant (ps ≥ .652).

To determine if the faster response times and decrease in switching cost were due to language factors or repetition of the task, the same analyses were completed comparing the Interpretation and Translation students together to the Monolingual controls. As with the previous RT analysis, there were main effects of phase (F(1,36) = 31.612, p < .001, η_p^2 = .468) and trial type (F(1,36) = 82.007, p < .001, η_p^2 = .695) and an interaction between these factors (F(1,36) = 8.763, p = .005, η_p^2 = .196). All other effects were non-significant (ps ≥ .260).

3.3.5.2 Mixing costs. Mixing costs represent the difference in performance on repetition trials in the mixed-task block and trials in the single-task blocks. The two groups did not differ in accuracy on either trial type at either phase (ps ≥ .414). Across the groups, accuracy was higher to single-task trials than to repetition trials in Phase 1 (Wilcoxon T = 201, Z = 2.420, p = .016), but there was no difference in Phase 2

($p = .123$). A three-way ANOVA on RTs with phase (first, second) and trial type (single, repetition) as within-subjects factors and group (Interpretation students, Translation students) as a between-subjects factor was computed. Main effects of phase ($F(1,25) = 39.302$, $p < .001$, $\eta_p^2 = .611$) and trial type ($F(1,25) = 103.076$, $p < .001$, $\eta_p^2 = .805$) were revealed. These were due to faster responses at Phase 2 and to single-task trials. These two factors also interacted ($F(1,25) = 43.403$, $p < .001$, $\eta_p^2 = .635$; Figure 6), revealing a decrease in mixing cost between Phase 1 and Phase 2. No effects with group were significant (ps $\geq .652$).

Table 10: Task-switching paradigm: mean response times and accuracy by group and phase

	Interpretation students		Translation students		Monolingual controls	
	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)	Response Time (ms)	Accuracy (%)
Phase 1						
Single task	465 (113)	98.8 (1.7)	493 (92)	99.2 (1.2)	475 (60)	98.6 (1.5)
Repetition	779 (222)	97.7 (2.1)	811 (222)	98.0 (2.3)	733 (93)	97.7 (2.5)
Switch	955 (310)	95.5 (3.7)	990 (270)	95.1 (3.8)	893 (128)	95.0 (4.0)
Mixing cost	314 (132	1.1 (2.6)	318 (155)	1.2 (1.5)	258 (116)	0.9 (2.5)
Switching cost	177 (125)	2.2 (3.4)	179 (95)	2.9 (2.6)	160 (70)	2.7 (3.2)
Phase 2						
Single task	409 (83)	97.7 (3.0)	440 (71)	97.9 (2.6)	419 (63)	97.6 (2.2)
Repetition	626 (199)	97.1 (2.4)	640 (196)	96.4 (4.7)	626 (130)	97.0 (2.3)
Switch	783 (295)	94.5 (4.1)	787 (253)	89.4 (7.7)	731 (154)	93.3 (5.2)
Mixing cost	217 (128)	0.6 (3.0)	200 (136)	1.5 (2.6)	206 (86)	0.6 (1.6)
Switching cost	158 (117)	2.6 (3.0)	147 (98)	7.0 (5.9)	106 (49)	3.8 (4.6)

Note: Values in parentheses are standard deviations.

To determine if the faster response times and decrease in mixing cost were due to language factors or repetition of the task, the same analyses were completed comparing the Interpretation and Translation students together to the Monolingual controls. As with the previous RT analysis, there were main effects of phase ($F(1,36) = 37.243$, $p < .001$, $\eta_p^2 = .408$) and trial type ($F(1,36) = 148.832$, $p < .001$, $\eta_p^2 = .805$) and an interaction between these factors ($F(1,36) = 16.582$, $p < .001$,

η_p^2 = .315). However, the three-way interaction was not significant (p = .175). All other effects were non-significant (ps ≥ .420).

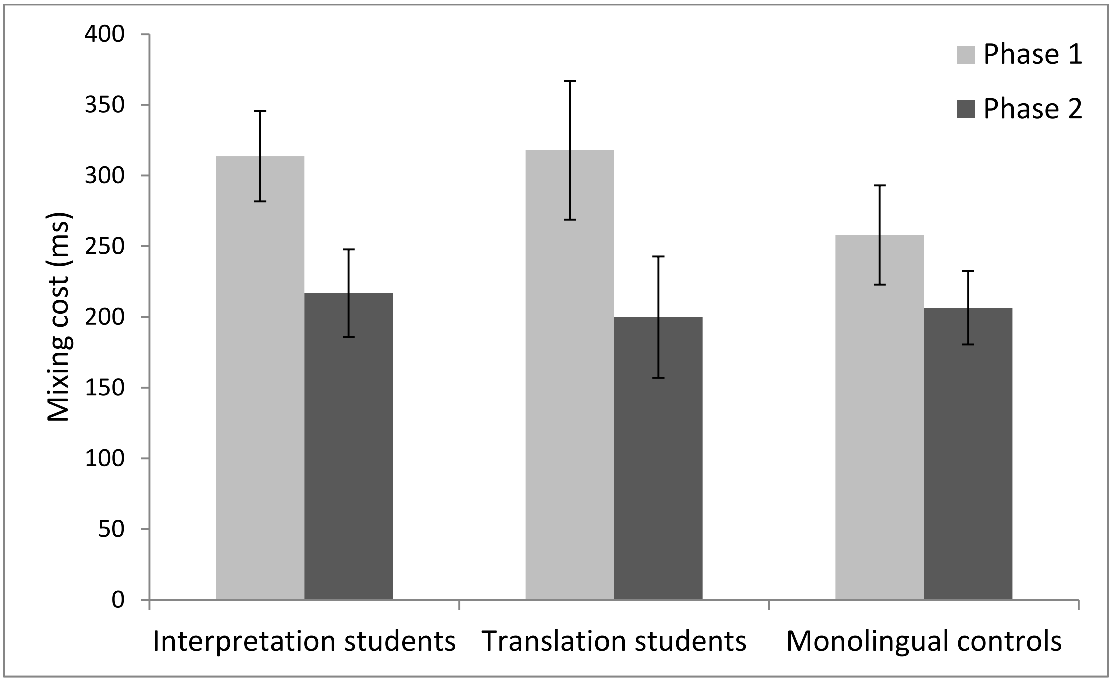

Figure 6: Mixing cost by group and phase. Error bars represent standard errors of the mean.

3.4 Discussion

3.4.1 Summary of the Main Results

This study investigated the effects of training in simultaneous interpretation on memory and executive functioning through a longitudinal design. The students of interpretation showed increases in verbal short-term and working memory, as well as spatial short-term memory, between Phase 1 and Phase 2. These changes appear to be specific to the students of interpretation. Additionally, these students showed an enhanced alerting effect in the ANT that was not visible in the two control groups.

Across the executive functioning tasks, a general speed-up in all groups was evident. In the color-word Stroop task, the speed-up was accompanied by a decrease in accuracy across the groups suggesting a change in priority between the phases. Further, all three groups showed diminished switching and mixing costs in the task-switching paradigm at Phase 2.

3.4.2 Evidence of Innate Abilities

Overall, the data suggest that there are not innate differences between interpreters and non-interpreters on the memory and executive functioning processes assessed here. Across the tasks, a main effect of group was apparent only on the conflict effect in the ANT. This difference in groups, however, was seen on the follow-up analysis comparing the two multilingual student groups to the monolingual controls. Thus, it represents an advantage for multilinguals in conflict resolution, replicating previous results in the bilingualism literature (e.g., Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Costa et al., 2008; Luk, De Sa, & Bialystok, 2011).

Looking beyond the memory and executive functioning tasks, the students of interpretation did distinguish themselves in terms of verbal intelligence, scoring higher than both the students of translation and the monolingual controls. While the difference between the interpretation students and monolingual controls is unsurprising given the formers' education in languages, the difference between the interpretation and translation students is telling. The degree required to enter the Master's programs in both Translation and Conference Interpreting is a Triennale (equivalent to a Bachelor's degree) in Languages. Thus, the two groups of students received the same education up to the point of entering a Master's program. It appears then that high verbal intelligence may be a necessary characteristic for simultaneous interpretation, or at least acceptance into a training program. To our knowledge this is the first report of higher verbal intelligence among interpreters. This is not, however, due to a lack of differences in previous studies, but rather the absence of verbal intelligence assessments. Future studies may consider including measures of verbal intelligence to examine this effect further.

It is important to note that this difference in verbal IQ represents a potentially confounding factor in studies of interpreters. In the present results, however, it appears not to have influenced other measures. No other differences across the phases were apparent. Further, it is unlikely that a difference in verbal IQ that was already present at Phase 1 would cause an improvement over time on other measures of cognition when the two groups began at equal levels.

3.4.3 Evidence of Trained Abilities

The present findings suggest that specific abilities are enhanced through training in simultaneous interpretation. In particular, the students of interpretation showed increases in verbal and spatial short-term memory and verbal working memory which were not as evident in the students of translation. Notably, these enhancements mirror the advantages in memory seen in the professional interpreters discussed in the previous chapter. It appears that these differences in memory processes are due to training in simultaneous interpretation, rather than due to innate characteristics. As mentioned in the previous chapter, verbal short-term and working memory are likely employed during SI to store content and rehearse pre-output translations. Thus, simultaneous interpretation may constitute targeted memory training which leaves the individual with increased memory span. The influence of SI on spatial short-term memory is less transparent. As put forth in the previous chapter, improved spatial short-term memory may be connected to the use of a meaning-based strategy in which interpreters retain the input at a non-verbal conceptual level (Fabbro & Gran, 1994). These two studies represent the first examination of spatial memory in interpreters; therefore replication of these findings is necessary. Additionally, the current study is only the second study to investigate verbal memory longitudinally. However, the present results are in discord with the previous longitudinal study (Macnamara & Conway, 2013). That study found no improvement in reading span and operation span in students of ASL interpretation. As mentioned above, though, this may be due to differences between unimodal and bimodal simultaneous interpretation. Interestingly, that study did not include measures of non-verbal memory, which may be relied upon more in ASL interpretation.

In addition to an increase in memory span, the Interpretation students also improved their performance on the intervening task in the tests of working memory, while the performance of the two control groups declined. This difference could represent a change in strategy, shifting more focus to the intervening task than the items to be recalled among the Interpretation students. However, in that case, a concomitant decrement in memory would be expected which is not present in the data. Alternatively, the improvement on both task components could arise from the increase in memory such that ceiling performance is reached on the item recall component, freeing up resources for the intervening task. This explanation could also accommodate the absence of an interpreter advantage on the intervening task in the previous chapter. Supposing that age effects and SI training effects are additive, the students would profit from both to reach ceiling performance, while the older professional interpreters only from SI training. However, this account ultimately fails on two points. First, the pattern of improvement on the intervening task was seen across the domains, while the improvement in item recall was only present in the verbal domain. Second, this explanation supposes that ceiling performance had been reached on item recall, which was not the case for either test of working memory. A final account of this finding relies on viewing the tests of working memory as dual task situations. Simultaneous interpretation may also be considered a dual task situation; thus the improvement on the intervening task may be due to training in managing two simultaneous tasks. This explanation gains support from a study showing an interpreter advantage over monolinguals in a verbal recall-tracking dual task paradigm (F. Padilla, Bajo, & Macizo, 2005). Further, given the interpreter advantage in mixing cost on a non-linguistic task-switching paradigm, it is likely that language management in SI relies, at least partially, on domain-general mechanisms, which could explain the cross domain improvement. Additional studies using dual task paradigms are needed to confirm this explanation.

A final effect of training in simultaneous interpretation was evident in the alerting effect of the ANT. The Interpretation students increased their alerting effect between Phase 1 and Phase 2, indicating that they made more use of cues which signaled the immediate appearance of the stimulus at Phase 2. This change could be due to the need to use and respond to cues during simultaneous interpretation. However, the change in alerting effect was small. Additionally, this advantage in the alerting network was not apparent among the professional interpreters examined in the previous chapter or a recent study (Morales et al., 2015). Thus, before accepting increased alertness as a signature of interpretation, more studies should confirm the result.

3.4.4 Evidence of Abilities Enhanced by SI Experience

A general increase in performance across the three groups was seen in response time on all tasks and in the switching and mixing costs in the task-switching paradigm. Among these, the mixing cost is of particular interest as professional interpreters distinguished themselves from multilinguals on this measure. The lack of an advantage for the students of interpretation in mixing cost may indicate that the mechanisms underlying sustained control are enhanced through experience with interpretation rather than the two-year training period. Alternatively, a case may be made that all three student groups are at peak performance and therefore any group differences are obscured. A second look at the data reveals that the Interpretation and Translation students had slightly higher mixing costs at Phase 1 than the Monolingual controls. At Phase 2, however, all three groups show roughly equivalent mixing costs around 200 ms, indicating that the decrease in mixing cost was smaller for the Monolingual controls. The 200 ms value may represent the maximum performance possible using our specific task conditions, and therefore the Monolingual controls could not improve beyond this point. These two accounts of the collective data could be disentangled by employing a task-switching paradigm that increases demands during the mixed-task block, potentially by increasing the number of task components (Steinhauser & Hübner, 2005).

3.4.5 Comparisons to Professional Interpreters

Attentive readers may have noticed that the students in the current study consistently showed better performance than the professional interpreters in the previous chapter, even at Phase 1. As mentioned above, the participants in the previous chapter were older than the students in the present study. Thus, the lower performance is likely due to age-related declines, which have been previously shown for both mixing costs (Kray & Lindenberger, 2000) and memory (Salthouse, 1994).

3.4.6 Limitations and Future Directions

The present study represents an initial effort to understand the evolution of cognitive processes in individuals learning to perform simultaneous interpretation. This study, however, does have some limitations which should be addressed in future projects. Primary among these limitations is the small sample size. With such a reduced size, it is possible that the results of a few individuals have swayed the general outcomes. A small sample size, though, may be inherent to studying this population, as a limited number of students enter training programs each year. Thus, future projects would benefit from recruiting participants from multiple training programs or multiple entry years in a single program. Further, collaborations with training programs would assist in guaranteeing the necessary number of students.

Such collaborations could be additionally beneficial to both sides. Researchers would gain much from understanding the training and selection procedures, as well as the specific nature of simultaneous interpretation. This knowledge would inevitably lead to better task selection. In the present study we examined areas of cognition that have previously been linked to enhancements in either interpreters or multilinguals generally. With a more intimate understanding of SI, though, additional areas of cognition benefitted by interpretation may be identified and investigated. Training programs, on the other hand, may benefit from collaborations through the modification and improvement of their selection and training procedures. By understanding which cognitive profiles and training methods create successful interpreters, their procedures may be refined.

Collaborations with training programs may address another shortcoming of the present study. We tested participants in the first three months of their Master's program, rather than before courses began, thus our participants may have already experienced some changes before our baseline testing. Completing the first test session before courses begin would potentially lead to a greater difference between test sessions.

An additional limitation of the present study is that we considered only the two-year training period and not later experience in the profession. Professional experience may lead to both quantitative and qualitative changes. In fact, the incongruous results in mixing cost between the present study and the study in the previous chapter may be related to professional experience. Future studies following professional interpreters at the start of their careers would greatly add to our understanding of the separate roles of SI training and SI experience.

A final comment pertains to the use of a monolingual control group. While this group added useful information about the effects of task repetition, they may not provide a fair comparison for differences in innate characteristics. Within any population individual differences are seen across cognitive abilities, some of which may be useful to the practice of simultaneous interpretation. Thus, multilinguals who excel in those abilities may be drawn to simultaneous interpretation; it is this effect that creates the differences seen between the students of interpretation and those of translation. The same spectrum of individual differences is present in monolinguals; however, due to their lack of proficiency in other languages, individuals who excel in "simultaneous interpretation abilities" will not pursue interpretation as a career path. Thus, innate differences between interpreters and some monolinguals may be more difficult to detect because language ability, not cognitive ability, provided the barrier to the monolinguals.

3.5 Concluding Remarks

The present study suggests that success in simultaneous interpretation is due to a mixture of innate and trained abilities. Verbal intelligence emerges as an ability which allows entrance into the field of simultaneous interpretation. Memory, on the other hand, which constitutes a commonly found interpreter advantage, is developed through training in simultaneous interpretation. Understanding which components of simultaneous interpretation are trainable and which are innate will allow for better selection and training programs moving forward. Future studies which examine students of interpretation will further contribute to the betterment of these programs.

CHAPTER 4

DOES SIMULTANEOUS INTERPRETATION TRAINING LEAD TO MORE BRAIN OR BETTER BRAIN?

4.1 Introduction

Santiago Ramón y Cajal once wrote “any man could, if he were so inclined, be the sculptor of his own brain” (Ramón y Cajal, 1999, p. xv). While Ramón y Cajal may not have intended this in a biological sense, research in the past two decades has shown that particular experiences do sculpt our brains at a biological level. Skills such as juggling, driving a taxi, playing the piano, and learning a language all leave an impression on the structure of the brain (e.g., Draganski et al., 2004; Imfeld, Oechslin, Meyer, Loenneker, & Jancke, 2009; Maguire et al., 2000; Mårtensson et al., 2012). Thus, it stands to reason that simultaneous interpretation, which is a complex and demanding skill, should also sculpt the brain.

To the best of our knowledge only two studies to date have examined the structural differences between professional simultaneous interpreters and non-interpreter multilinguals. A study focused on differences in white matter found that interpreters had smaller fractional anisotropy values (FA; a measure of the directedness of diffusion which reflects white matter integrity in the brain) in the left anterior insula, right inferior parietal lobe, right dorsal caudate, and areas of the corpus callosum (Elmer, Hänggi, Meyer, & Jäncke, 2011). In another study, Elmer and colleagues investigated the differences in gray matter structure between these same groups (Elmer, Hänggi, & Jäncke, 2014). The interpreters showed smaller gray matter volumes than the multilinguals in several regions which support language control including left supramarginal gyrus and bilateral inferior frontal gyrus. Additionally, that study revealed negative correlations between the number of hours of interpretation experience and the gray matter volume in a subset of these regions (after controlling for age differences). The authors speculated that the smaller gray matter volume may be due to cortical pruning that improves functional specialization and efficiency.

As noted in the previous chapter, and by Perani (2005) in specific relation to biological differences, it is difficult to understand if group differences are due to innate characteristics or rather are the consequence of training. Elmer and colleagues partially addressed this issue through the correlational analyses with hours of interpretation experience (Elmer et al., 2014). They posited that the differences in areas which showed a correlation with SI experience were training-related, while the areas that showed group differences but not a correlation were persistent group differences. This dissociation, however, did not consider the separate role that initial training in simultaneous interpretation may have on brain structure. A discrete role of training is supported by the training-related changes in memory seen in the previous chapter. Thus, the regions where differences were attributed to persistent differences may, in fact, have been changed by initial training in SI. To comprehend the full picture of SI influence on the brain, changes occurring during the training period need to be examined.

While no published studies have examined the structural changes associated with training in simultaneous interpretation, several studies have investigated changes related to language learning. The dominant trend among these studies is an increase in gray matter density and white matter integrity with increased language proficiency (see Li, Legault, & Litcofsky, 2014 for a review). This pattern suggests that increased language proficiency is associated with greater amassment of brain volume, in contrast to the greater efficiency seen in professional interpreters. As training in simultaneous interpretation involves aspects of language learning as well as the acquisition of the skill set used by professional interpreters, it appears to sit at the crossroad between greater amassment and greater efficiency.

To examine the influence that training in simultaneous interpretation has on brain structure, we acquired neuroimaging data from students earning a Master in Conference Interpreting at the beginning and end of their two-year program. Additionally, to isolate the specific effect of training in SI, as opposed to natural changes over time or changes due to language practice, we also collected longitudinal data from students earning a Master in Translation. These students represent the ideal comparison to isolate the effects of training in SI as they also speak multiple languages and were engaged in language training during the intervening two years. Critically, though, the students of translation had no experience with simultaneous interpretation. Similar to the previous chapter, the use of these two groups longitudinally allowed us to disentangle innate differences between the groups from differences specifically due to SI training.

The two groups completed neuroimaging sessions at the start and finish of their Master's degrees. Gray matter was assessed using T1-weighted anatomical imaging and voxel-based morphometry (VBM) analyses. White matter was examined with diffusion tensor imaging (DTI) and virtual *in vivo* dissections after whole brain tractography. Based on the studies presented above by Elmer and colleagues, we may expect smaller gray matter volumes and less white matter integrity among the students of interpretation than the students of translation. The appearance of such a difference across the phases would suggest an innate difference between the groups. If the difference appears only after training, however, it may be specifically related to SI training. Alternatively, we may expect increases in gray matter volume and white matter integrity similar to those seen in language learning.

4.2 Methods

4.2.1 Participants

The participants in the current study were a subset of those examined in the previous chapter. Fifteen students of interpretation (11 females) and eight students of translation (6 females) completed the neuroimaging session at both time points. As in the previous chapter the students were tested at the start of their Master's program and again approximately two years later at the conclusion of the program. All participants were right-handed and had no known neurological or psychiatric problems. Participants gave written informed consent and were compensated for their time. The study was approved by the ethical committee of "Istituto IRCCS E. Medea – La Nostra Famiglia."

4.2.2 T1 and DTI Acquisition

All images were acquired on a 3-Tesla Philips Achieva whole-body scanner at the Santa Maria della Misericordia Hospital in Udine, Italy. T1-weighted images were acquired using an 8-channel head coil and a 3D ultrafast gradient echo sequence (190 axial slices, TR = 8.2 ms, TE = 3.7 ms, 1 x 1 x 1 mm voxel resolution, acquisition matrix = 240 x 240). DTI acquisition proceeded using 57 contiguous axial slices (1.875 x 1.875 x 2.1 mm voxel resolution). For each slice, 64 diffusion-weighted images (b = 1000 s/mm^2) and one non-weighted image were acquired.

4.2.3 VBM Processing and Statistical Analyses

Voxel-based morphometry (VBM) analyses were conducted using the VBM8 toolbox (http://dbm.neuro.uni-jena.de/) in SPM8 (Statistical Parameter Mapping; Wellcome Department of Cognitive Neurology, UCL, London, UK). Preprocessing began by manually setting the coordinate origin of each image to the anterior commissure and adjusting for large rotations of the head. The images were then segmented, spatially normalized, and modulated in a single step. Segmentation used SPM8's "new segmentation approach." Spatial normalization was completed using the diffeomorphic anatomical registration using exponentiated lie algebra (DARTEL) technique. Modulation used non-linear terms. The resulting gray matter images were spatially smoothed with an 8 mm full-width-at-half-maximum Gaussian filter.

Group differences were analyzed using a 2-sample *t*-test on the smoothed images across the time points. Contrasts with Interpretation students vs. Translation students and vice versa were extracted. To analyze the effects of phase and the interaction of phase and group a "flexible factorial" design was used with the factors participant, phase, and group. The main effect of phase was examined using two contrasts (Phase 1 vs. Phase 2 and Phase 2 vs. Phase 1). Additionally, to explore the interaction between phase and group two contrasts were extracted. The first assessed increases in volume between Phase 1 and Phase 2 that were greater in the Interpretation students than Translation students (Interpretation students Phase 1: -1; Interpretation students Phase 2: +1; Translation students Phase 1: +1; Translation students Phase 2: -1). This contrast can alternatively be viewed as decreases that were greater in the Translation students than Interpretation students (examination of the volume estimates will elucidate this issue). The second contrast evaluated increases in volume that were greater for the Translation students (alternatively, decreases that were greater for the Interpretation students; Interpretation students Phase 1: +1; Interpretation students Phase 2: -1; Translation students Phase 1: -1; Translation students Phase 2: +1). For all contrasts threshold masking was employed with an absolute threshold of 0.1. Data were generated at $p < .001$ uncorrected and then clusters were selected using cluster-wise $p < 0.05$, corrected for multiple comparisons using a Family-wise Error (FWE) correction. The Brodmann and automated anatomical labeling (AAL) templates in MRIcroN (http://www.mccauslandcenter.sc.edu/mricro/mricron/) were used to find the likely Brodmann area (BA) and anatomical region for each cluster.

4.2.4 DTI Processing, Tract Dissection, and Statistical Analyses

The DTI data were processed using ExploreDTI (http://www.exploredti.com). Subject motion and geometric distortions induced by eddy currents were corrected in a single step through rotation of the b-matrix. A b-spline interpolated streamline algorithm was then used to perform whole brain tractography (stepsize: 1 mm; FA threshold: 0.2; angle threshold: 35°). Using software written by Natbrainlab (Thiebaut de Schotten et al., 2011), the results of the tractography were imported to TrackVis (http://www.trackvis.org), where virtual *in vivo* dissections were performed.

Five tracts of interest were selected for dissection based on their known connection to language (Figure 7). The importance of the arcuate fasciculus for language functions has been recognized over a century (Geschwind, 1967). Originally, this pathway was conceived of as a direct connection between Broca's area and Wernicke's area. More recently, however, an indirect pathway mediated through the inferior parietal lobule has been additionally identified in both post mortem and virtual *in vivo* dissections (Catani, Jones, & ffytche, 2005). Thus, it is now accepted that the arcuate fasciculus is composed of three segments. The classic connection between Broca's area and Wernicke's area is termed the long segment. The connection between Broca's area and the inferior parietal lobule is the anterior segment, while the tract connecting Wernicke's area to the inferior parietal lobule is the posterior segment. All three segments of the arcuate fasciculus were dissected in the present study. Additionally, the frontal aslant tract, which has been recently identified and connects the inferior frontal gyrus and the superior frontal gyrus, was included (Catani et al., 2012). This tract has been linked to verbal fluency and thus may be particularly involved in simultaneous interpretation (Catani et al., 2013). The final tract selected was the uncinate fasciculus which connects the anterior temporal lobe and the orbitofrontal region (Catani, Howard, Pajevic, & Jones, 2002). This tract was included due to its connection with lexical retrieval and semantic associations (Catani & Mesulam, 2008; Lu et al., 2002).

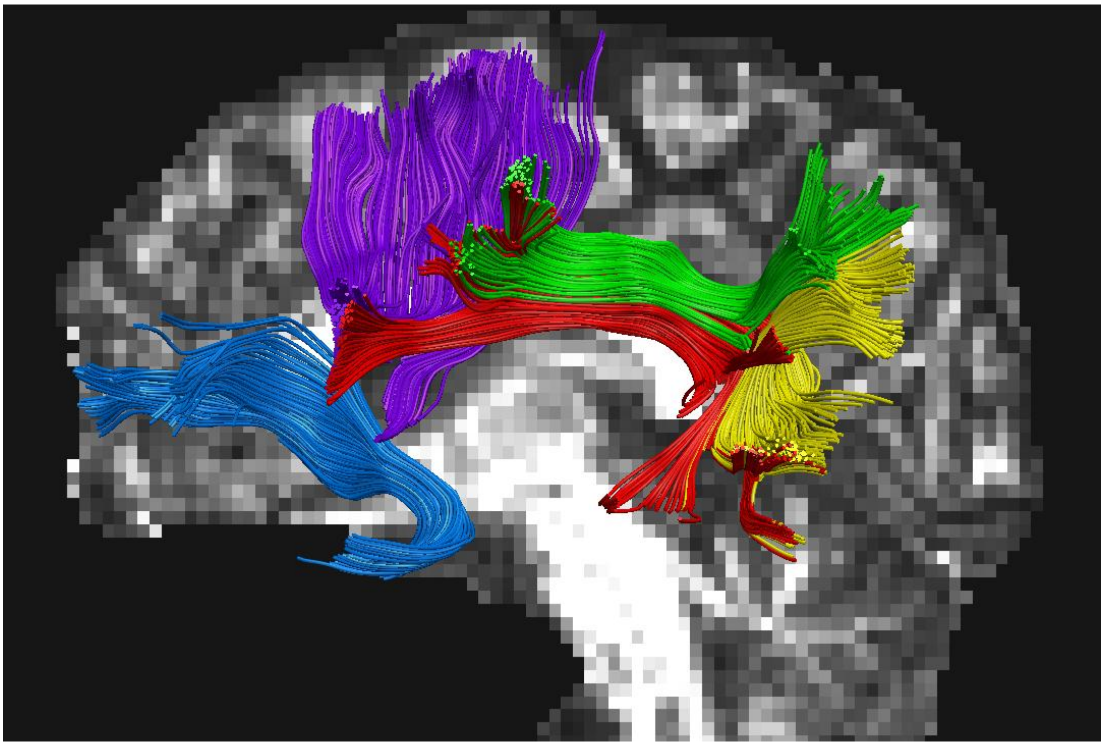

Figure 7. The five white matter tracts dissected for each participant. Shown here in the left hemisphere for one participant overlaid on the FA map. Anterior segment of the arcuate fasciculus – green; posterior segment of the arcuate fasciculus – yellow; long segment of the arcuate fasciculus – red; frontal aslant tract – purple; uncinate fasciculus – blue.

These five tracts of interest were dissected bilaterally in the two hemispheres. Dissection of each tract proceeded by manually defining two regions of interest on the fractional anisotropy images in the three planes and using these as seed regions for tracking. The demarcation of the regions of interest followed previous studies of tractography. The three segments of the arcuate fasciculus were dissected using three regions of interest (Catani, Jones, & ffytche, 2005; Catani & Mesulam, 2008). The first region was Broca's territory, encompassing the inferior frontal gyrus, middle frontal gyrus, and precentral gyrus. The second region was Wernicke's territory including the middle and superior temporal gyri. The third region was Geschwind's territory composed of the supramarginal and angular gyri. The anterior segment was tracked using Broca's and Geschwind's territories, the posterior segment using Wernicke's and Geschwind's, and the long segment using Broca's and Wernicke's. The frontal aslant tract was dissected using regions of interest in the inferior frontal gyrus and superior frontal gyrus (Catani et al., 2012). Finally, the regions of interest used to track the uncinate fasciculus were the white matter of the anterior temporal lobe and the white matter of the anterior floor of the extreme capsule (Catani & Thiebaut de Schotten, 2008).

For each tract, three measures of interest were extracted. The first measure was the total volume of the tract. Second, the fractional anisotropy (FA) was extracted. FA values range between 0 and 1 and indicate the degree of anisotropy in a diffusion process. Values near 0 represent isotropic diffusion, while a value of 1 indicates diffusion along only one axis. In diffusion imaging, FA is taken to reflect fiber density, axonal diameter, and myelination of the white matter. The final measure considered was the radial diffusivity (RD). This measures the diffusivity in the directions orthogonal to the eigenvector of the diffusion tensor. Small values of RD may indicate greater myelination, larger axonal diameter, or greater fiber density. These three measures were analyzed using mixed effects two-way ANOVAs with phase (first, second) as a within-subjects factor and group (Interpretation students, Translation students) as a between-subjects factor. Analyses were computed separately for each tract of interest in each hemisphere.

4.3 Results

4.3.1 VBM

The comparisons examining the main effect of phase revealed areas of volumetric changes in both directions. Across the groups a decrease in volume was seen in large areas including bilateral subcortical regions, bilateral parietal regions, right cingulum, and left cerebellum (Table 11). The opposite comparison revealed volumetric increases in different areas of the left cerebellum (Table 11).

Table 11: Areas showing changes in volume between the phases

Anatomical Localization	BA	MNI Coordinates x	y	z	Cluster *p*-corr.	Peak Z-value	Voxels per cluster
Phase 1 vs. Phase 2							
L. putamen	25	-14	9	-6	< .0001	6.49	28293
R. insula*	48	38	-6	-8		5.64	
R. insula	48	39	4	-3		5.43	
R. middle cingulum		10	-21	34	< .0001	5.39	3318
R. middle cingulum		10	-36	42		4.57	
R. middle cingulum		10	-10	48		3.95	
L. cerebellum 8		-9	-67	-30	.033	5.21	587
L. cerebellum 6		-16	-64	-29		4.55	
L. cerebellum 8*		-32	-52	-38		3.54	
R. precuneus	7*	12	-57	40	.026	4.63	620
R. precuneus	23*	14	-57	28		4.33	
L. precuneus	7*	2	-67	31		3.35	
L. precuneus		-14	-45	40	.003	4.35	958
L. middle cingulum*		-16	-33	36		4.03	
L. precuneus	23*	-12	-54	36		3.75	
L. postcentral gyrus	4	-48	-18	43	.005	4.26	902
L. rolandic operculum*	48	-46	-19	25		3.57	
L. supramarginal gyrus	2	-58	-28	40		6.26	
Phase 2 vs. Phase 1							
L. cerebellar crus2*		-2	-81	-41	< .0001	4.44	3224
L. cerebellar crus1		-48	-76	-36		3.86	
L. cerebellar crus2		-24	-82	-38		3.57	
L. cerebellum 8		-27	-37	-53	0.014	5.02	716
L. cerebellum 8*		-24	-48	-62		4.31	
L. cerebellum 10		-20	-33	-45		3.98	

*Localization or BA based on the closest labeled voxel within 8 mm.

Comparisons between the groups across the phases showed no regions of larger volume in the Interpretation students than the Translation students. The opposite comparison, however, signaled greater volume in the Translation students in several regions of the left temporal lobe and bilateral hippocampal regions (Figure 8 and Table 12).

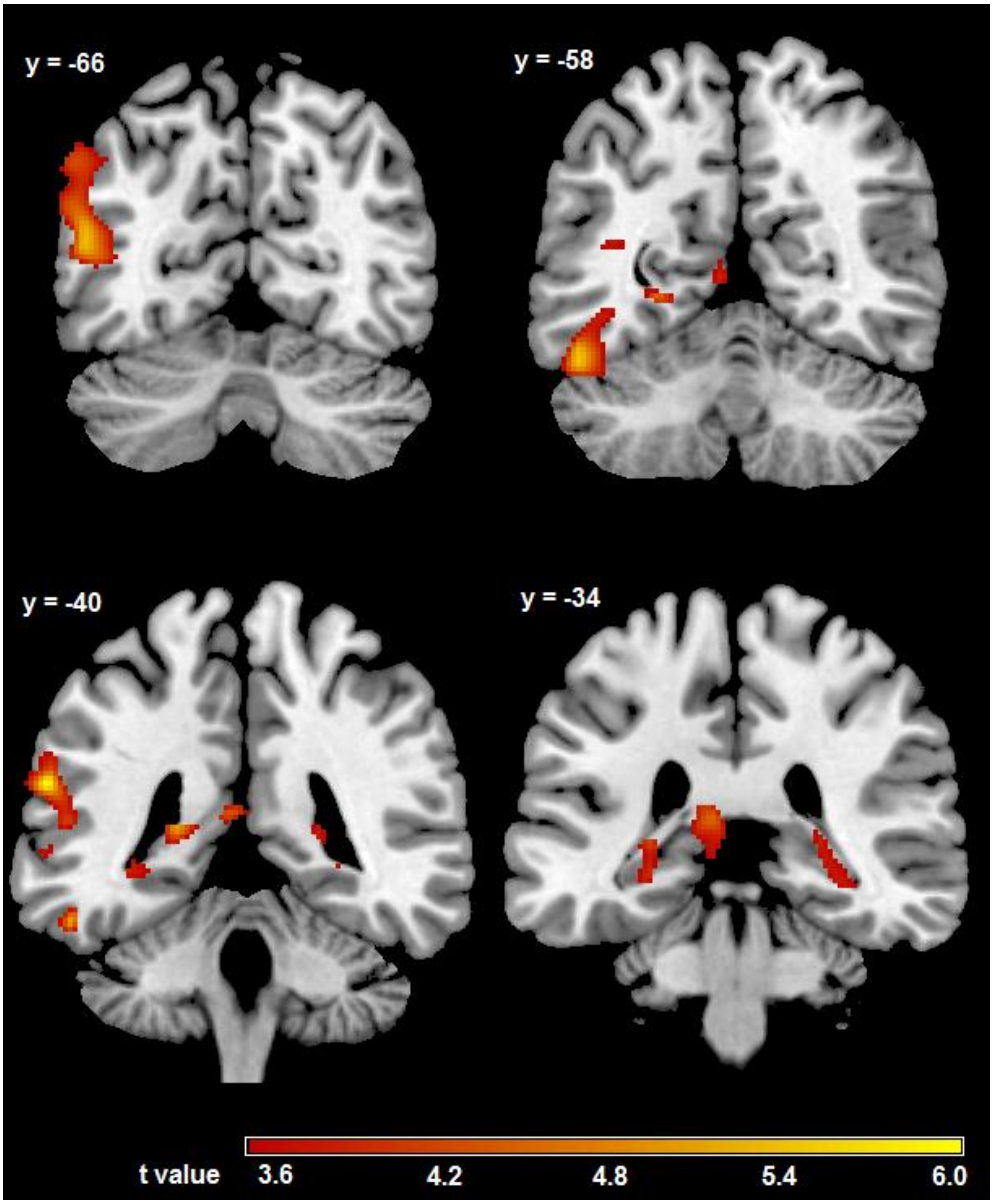

Figure 8: Areas showing larger volume in Translation students than Interpretation students.

Table 12: Areas showing larger volume in Translation students than Interpretation students

Anatomical Localization	BA	MNI Coordinates x	y	z	Cluster *p*-corr.	Peak Z-value	Voxels per cluster
L. superior temporal gyrus	42	-62	-40	22	0.033	5.14	728
L. middle temporal gyrus	22	-56	-37	9		4.05	
L. superior temporal gyrus*	22*	-68	-48	22		3.81	
L. inferior temporal gyrus	20	-54	-42	-20	0.009	5.12	1005
L. inferior temporal gyrus	37	-50	-55	-21		4.91	
L. inferior temporal gyrus*	20	-64	-46	-23		4.40	
L. middle temporal gyrus	39	-50	-64	18	0.006	4.76	1105
L. angular gyrus*	39*	-56	-72	28		4.68	
L. angular gyrus*	39*	-54	-67	40		3.98	
L. precuneus*	27	-21	-43	7	< .0001	4.66	1814
L. lingual gyrus	19	-26	-54	-2		4.45	
L. precuneus*	29*	-3	-43	13		4.23	
R. hippocampus	20	32	-27	9	0.044	3.83	669
R. hippocampus*	37	28	-28	-2		3.66	
R. lingual gyrus*	37*	30	-45	0		3.64	

*Localization or BA based on the closest labeled voxel within 8 mm.

Two final planned comparisons investigated the interaction of phase and group in changes in gray matter volume. The examination of larger decreases in volume in the Interpretation students than the Translation students yielded no areas of significant difference. The reverse comparison, which considered larger decreases among the Translation students than the Interpretation students, however, resulted in several areas of difference (examination of the volume estimates indicated that the interaction was due to larger decreases among the Translation students; Figure 9 and Table 13). These included a large region in the left hemisphere which had peaks in the putamen and hippocampus and extended to the amygdala, fusiform gyrus, and middle temporal pole. Similarly, in the right hemisphere a cluster was seen in the putamen which extended to the amygdala. Additional clusters were seen in the left superior temporal lobe, left cuneus/superior occipital lobe, right thalamus, and right calcarine region.

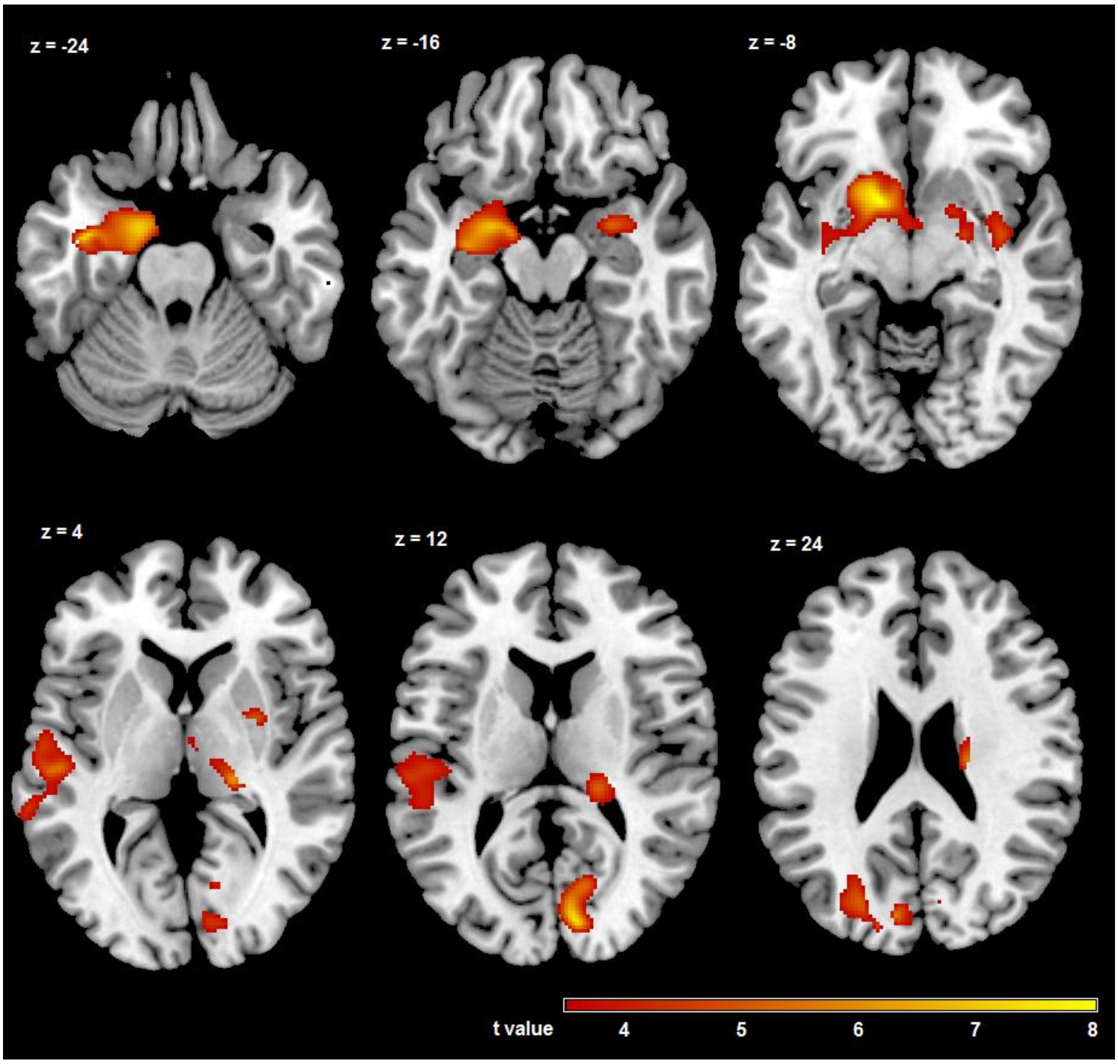

Figure 9: Areas showing smaller decreases in Interpretation students than Translation students.

Table 13: Areas showing a smaller decrease in Interpretation students than Translation students

Anatomical Localization	BA	MNI Coordinates x	y	z	Cluster *p*-corr.	Peak Z-value	Voxels per cluster
L. putamen	25	-12	8	-8	< .0001	5.38	4430
L. hippocampus	28	-15	-3	-21		5.17	
L. hippocampus	20	-30	-9	-20		5.07	
R. calcarine sulcus	17	10	-84	13	0.001	5.29	1150
R. calcarine sulcus	17	8	-73	16		5.18	
R. calcarine sulcus	18	12	-70	7		4.24	
R. thalamus	27*	18	-25	1	0.006	4.80	849
R. thalamus*		14	-28	18		4.41	
R. caudate*	48*	21	-18	24		4.34	
L. cuneus	19	-9	-82	24	< .0001	4.34	855
L. superior occipital lobe	19	-24	-76	24		4.08	
L. superior occipital lobe	18	-16	-87	24		3.71	
L. superior temporal lobe	48	-52	-21	4	< .0001	4.25	1400
L. superior temporal lobe	48	-58	-13	7		4.16	
L. supramarginal gyrus	48	-54	-24	15		3.72	
R. amygdala	34	28	-3	-15	0.004	4.19	938
R. putamen	48*	28	0	6		4.04	
R. insula*	48	38	-6	-8		3.93	

*Localization or BA based on the closest labeled voxel within 8 mm.

4.3.2 DTI

4.3.2.1 Left hemisphere.

4.3.2.1.1 Anterior segment of the arcuate fasciculus. The analysis on volume revealed a larger left anterior segment in the Interpretation students than the Translation students ($F(1,21) = 5.963$, $p = .024$, $\eta_p^2 = .221$). This group effect was not modulated by phase ($p = .759$), nor did phase have an effect across the groups ($p = .198$). The analysis on FA, however, did reveal a main effect of phase ($F(1,21) = 4.580$, $p = .044$, $\eta_p^2 = .179$), with FA decreasing between Phase 1 and Phase 2. While there was no main effect of group ($p = .126$), there was a marginal interaction between phase and group ($F(1,21) = 3.903$, $p = .061$, $\eta_p^2 = .157$). This interaction was due to a marginally larger decrease in FA between the phases among the Translation students than the Interpretation students ($t(21) = 1.976$, $p = .061$). An interaction between phase and group was also evident in the analysis on RD ($F(1,21) = 7.558$, $p = .012$, $\eta_p^2 = .265$). In this case the Interpretation students showed a decrease in RD across the phases ($t(14) = 2.688$, $p = .018$), while the Translation students showed a numerical increase ($p = .205$). There were no main effects of phase and group in the RD analysis (ps $\geq .355$).

4.3.2.1.2 Posterior segment of the arcuate fasciculus. The Interpretation students had a larger posterior segment in the left hemisphere than the Translation students ($F(1,21) = 9.088$, $p = .007$, $\eta_p^2 = .302$). No other effects in the analysis on volume were significant (ps $\geq .378$). Both phase and group showed main effects in the analysis on FA ($F(1,21) = 9.523$, $p = .006$, $\eta_p^2 = .312$ and $F(1,21) = 5.056$, $p = .035$, $\eta_p^2 = .194$, respectively). FA decreased between Phase 1 and Phase 2 and the Interpretation students had higher values than the Translation students. The interaction between these factors was not significant ($p = .935$). A main effect of group was also evident in the analysis on RD ($F(1,21) = 4.334$, $p = .050$, $\eta_p^2 = .171$), the Interpretation students showed smaller values than the Translation students. No other effects in the RD analysis were significant (ps $\geq .754$).

4.3.2.1.3 Long segment of the arcuate fasciculus. The long segment in the left hemisphere could not be tracked for one male Translation student at Phase 1 and two female Interpretation students at Phase 2; the data from these three participants were excluded from the present analyses. A main effect of group was seen in the analyses on volume ($F(1,18) = 8.613$, $p = .009$, $\eta_p^2 = .324$), FA ($F(1,18) = 12.249$, $p = .003$, $\eta_p^2 = .405$), and RD ($F(1,18) = 4.907$, $p = .040$, $\eta_p^2 = .214$). The Interpretation students showed larger volume and FA values, and a smaller RD value than the Translation students. Further, there was a marginal interaction of phase and group in the RD analysis (F(1,18) = 3.402, p = .082, η_p^2 = .159), due to a decrease in RD among the Interpretation students (t(12) = 3.740, p = .003). No other effects were significant (ps ≥ .143).

4.3.2.1.4 Frontal aslant tract. The frontal aslant tract in the left hemisphere could not be tracked for one male and one female Translation student at Phase 1, one female Translation student at either phase, and one female Interpretation student at Phase 2; the data from these four participants were excluded from the present analyses. No significant effects on any of the measures were revealed (ps ≥ .194).

4.3.2.1.5 Uncinate fasciculus. Main effects of group were seen in the volume (F(1,21) = 4.442, p = .047, η_p^2 = .175) and RD analyses (F(1,21) = 4.550, p = .045, η_p^2 = .178). These effects indicated a larger volume and smaller RD value among the Interpretation students than the Translation students. No other effects were significant (ps ≥ .213).

4.3.2.2 Right hemisphere.

4.3.2.2.1 Anterior segment of the arcuate fasciculus. The anterior segment in the right hemisphere could not be tracked at Phase 1 for one female Translation student; the data from this participant were excluded from the present analyses. The analysis on volume in the right hemisphere revealed a larger anterior segment for the Interpretation students (F(1,20) = 10.285, p = .004, η_p^2 = .340). Additionally, an interaction between phase and group was evident (F(1,20) = 7.050, p = .015, η_p^2 = .261), due to a significant decrease among the Interpretation students (t(14) = 3.637, p = .003) and a non-significant increase among the Translation students (p = .514). Further, in the FA analysis, main effects of phase, as a trend,

($F(1,20) = 4.280$, $p = .052$, $\eta_p^2 = .176$) and group ($F(1,20) = 13.366$, $p = .002$, $\eta_p^2 = .401$) were present. FA decreased between Phase 1 and Phase 2 and was larger in Interpretation students than Translation students. There was no interaction between these factors ($p = .962$). There were no significant effects in the analysis on RD (ps $\geq .134$)

4.3.2.2.2 Posterior segment of the arcuate fasciculus. There were no significant effects on any of the three measures in the right hemisphere (ps $\geq .175$).

4.3.2.2.3 Long segment of the arcuate fasciculus. This segment was trackable at both time points in only seven Interpretation students (6 females) and three Translation students (2 females). Previous research has shown that about 60% of individuals lack a long segment in the right hemisphere (Catani et al., 2007); thus, the percentage seen here is in line with the general population.[2] Given the small number of participants with this tract, non-parametric tests were used. No differences were apparent between the phases or groups (ps $\geq .142$).

4.3.2.2.4 Frontal aslant tract. The frontal aslant tract in the right hemisphere could not be tracked for one female Interpretation student at either phase; the data from this participant were excluded from the present analyses. A main effect of phase was revealed in the volume analysis ($F(1,20) = 4.423$, $p = .048$, $\eta_p^2 = .181$), indicating an increase in volume from Phase 1 to Phase 2. Additionally, there was an interaction of phase and group in the FA analysis ($F(1,20) = 4.346$, $p = .050$, $\eta_p^2 = .179$), due to a non-significant decrease in the Interpretation students ($p = .065$) and a non-significant

[2] The coherence between the groups tested here and the general population is potentially surprising. Previous research has shown a connection between greater bilaterality of the long segment and verbal memory (Catani et al., 2007). Additionally, the same study showed that about 70% of females do have a long segment in the right hemisphere. Thus, our participant groups may have been expected to show a higher rate of individuals with a right long segment given their focus on language and high number of female participants.

increase in the Translation students (p = .315). No other effects were significant (ps ≥ .123).

4.3.2.2.5 Uncinate fasciculus. The uncinate fasciculus could not be tracked for one female Translation student at Phase 1; the data from this participant were excluded from the present analyses. There were no significant effects on any of the three measures (ps ≥ .119).

4.3.2.3 Summary of DTI results. Figure 10 provides a graphical summary of these results.

4.3.2.3.1 Main effects of phase. Fractional anisotropy values decreased between Phase 1 and Phase 2 in the anterior segment of the arcuate fasciculus bilaterally, as well as in the left posterior segment of the arcuate fasciculus. Additionally, the volume of the right frontal aslant tract increased between Phase 1 and Phase 2.

4.3.2.3.2 Main effects of group. The Interpretation students had a stronger arcuate fasciculus in the left hemisphere than the Translation students. In the anterior segment this was seen as greater volume, and in both the posterior and long segments as greater volume, higher fractional anisotropy, and lower radial diffusivity. Additionally, the Interpretation students showed greater volume and higher fractional anisotropy in the anterior segment of the arcuate fasciculus in the right hemisphere. Finally, the Interpretation students had greater volume and lower radial diffusivity in the left uncinate fasciculus.

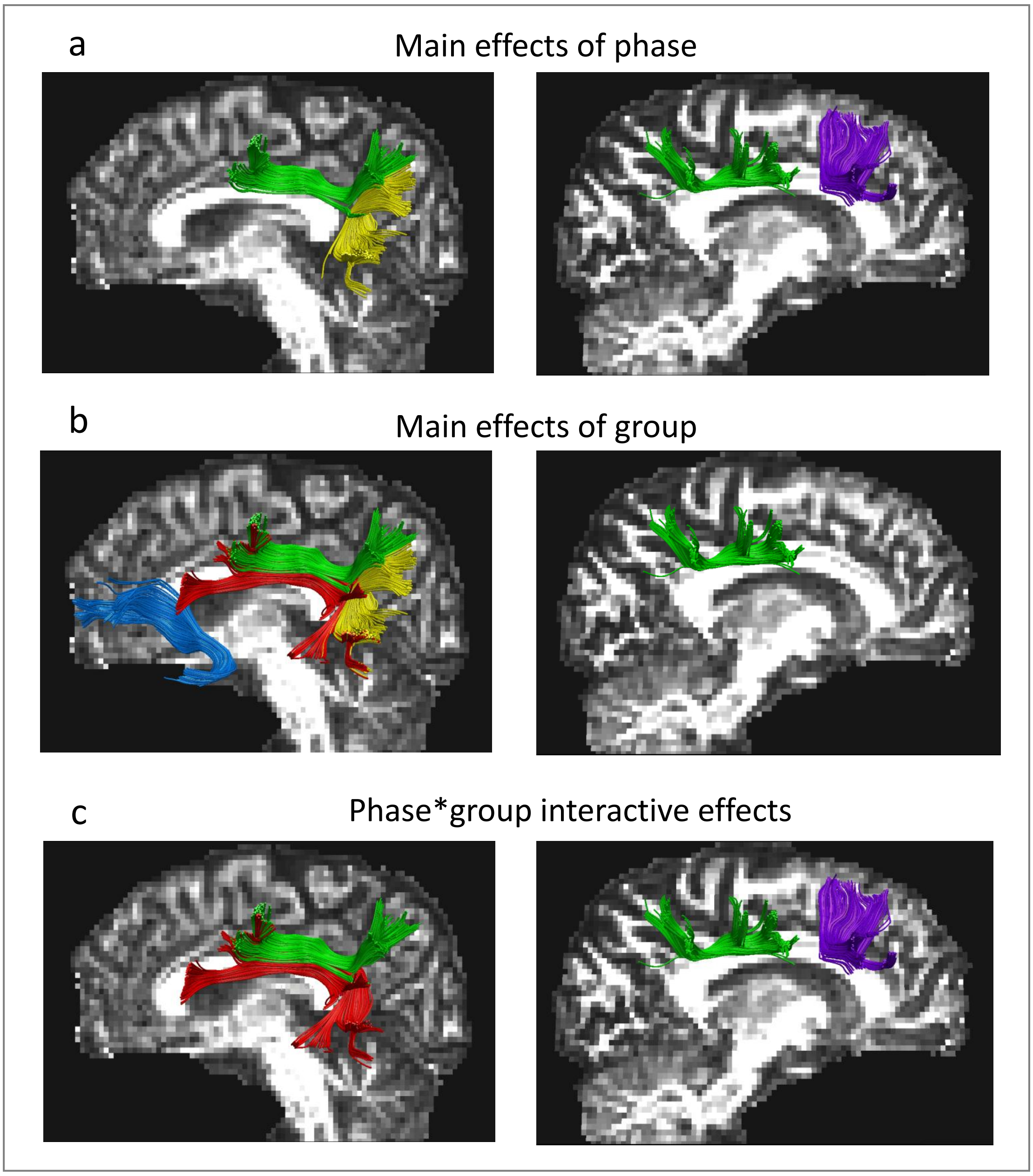

Figure 10: White matter tracts showing differences between (a) phases, (b) groups, or (c) an interaction between phase and group. See text for the specific measures showing differences.

4.3.2.3.3 Interactions of phase and group. Two interactive effects were seen in the anterior segment of the left arcuate fasciculus. The Translation students showed a larger decrease in fractional anisotropy than the Interpretation students.

Additionally, the Interpretation students showed a decrease in radial diffusivity, while the Translation students showed an increase. Further, in the long segment of the left arcuate fasciculus, the Interpretation students showed a decrease in radial diffusivity, while the Translation students showed no change. In the right hemisphere, the Interpretation students showed a decrease in volume of the anterior segment of the arcuate fasciculus, and the Translation students showed an increase. Finally, in fractional anisotropy of the right frontal aslant tract, the Interpretation students decreased and the Translation students increased.

4.4 Discussion

4.4.1 Summary of the Main Results

The present study investigated the changes in gray and white matter associated with learning to perform simultaneous interpretation using a longitudinal design. Changes between Phase 1 and Phase 2 in both gray and white matter were seen across the groups, with a general trend of decreased gray matter volume and specific decreases in white matter integrity. Across the phases, the students of interpretation showed less gray matter volume in left temporal and bilateral hippocampal regions and greater white matter integrity in the language connections, primarily in the left hemisphere. Effects related to training were also evident in both gray and white matter. The students of interpretation showed a smaller volumetric decrease in gray matter in bilateral subcortical structures, as well as in the left temporal lobe. Similarly, in the language tracts of the left hemisphere, the students of interpretation displayed less susceptibility to decreased white matter integrity.

4.4.2 General Longitudinal Effects

In both the gray matter and white matter analyses differences between Phase 1 and Phase 2 were seen across the groups. Large areas in bilateral subcortical and parietal regions, as well as a smaller region in the left cerebellum, showed a decrease in volume over time. This general decrease is in line with previous studies showing age-related declines in gray matter volume that are observable even in the early twenties (e.g., Ge et al., 2002; Walhovd et al., 2005). As these declines are generally seen across the cortex, the preservation of the frontal areas as well as the increase of volume in left cerebellar regions may be related to language learning effects on brain structure. Indeed, increased gray matter volume in the middle and inferior frontal gyri have been seen in connection with language learning (Mårtensson et al., 2012; Stein et al., 2012).

General age-related declines may also explain the decrease in fractional anisotropy seen in the arcuate fasciculus, specifically, in the anterior segment bilaterally and the left posterior segment. The change in the anterior segments is particularly supported by a study which found a negative correlation between age and FA for the tracts adjacent to the middle frontal gyrus bilaterally (Grieve, Williams, Paul, Clark, & Gordon, 2007). In addition to this decrease in FA, an increased volume was seen in the right frontal aslant tract. A previous study examining this tract exclusively in the left hemisphere found a correlation between verbal fluency and tract integrity (Catani et al., 2013). The present results may suggest a role in mediating verbal fluency for the frontal aslant tract in the right hemisphere as well. As verbal fluency may be particularly important in both interpretation and translation, growth in regions supporting this function may be expected with intense training. The right-lateralized nature of this effect is at first incongruous with the general left-lateralization of language, however, it may signal a shift towards bilateral language processing in these highly trained individuals. Such bilateral processing has been previously suggested in interpreters and highly proficient bilinguals (e.g., Darò, 1989; Gran & Fabbro, 1988).

4.4.3 Group Differences in Brain Structure

The Interpretation students had smaller gray matter volumes than the Translation students in left posterior temporal regions and bilateral hippocampus. These results could be viewed as evidence of either greater amassment among the Translation students or greater efficiency among the Interpretation students. Considering that these regions have been previously implicated in the systems used to learn and represent spoken words (Breitenstein et al., 2005; Davis & Gaskell, 2009), an advantage in efficiency among the Interpretation students seems probable. Spoken language is a cornerstone of simultaneous interpretation; additionally, the Interpretation students had higher verbal IQs and self-ratings in speaking at Phase 1 than the Translation students (see sections 3.2.2.1 and 3.2.2.2). These findings suggest that the Interpretation students had improved their efficiency in spoken language prior to beginning the Master's program. In fact, this ability may have contributed to their interest and/or acceptance into the program. It is interesting to note that while the professional interpreters in the study by Elmer and colleagues also showed decreased gray matter volume, these differences were not seen in the regions discussed here (Elmer et al., 2014). Both that study and the present study examined a small number of participants, thus, replication of both findings is needed to determine the consistency of this difference. If the results hold, however, the lack of differences between professional interpreters and multilinguals in these regions may be due to the eventual pruning of the areas used in spoken word learning among the multilinguals.

The students of interpretation additionally showed greater white matter integrity than the students of translation in all three segments of the left arcuate fasciculus, as well as in the right anterior segment and the left uncinate fasciculus. The left arcuate fasciculus has a well-known connection with language production and the uncinate fasciculus is thought to be involved in lexical retrieval and semantic associations (Catani & Mesulam, 2008; Lu et al., 2002). Thus, the higher integrity in these tracts may be related to the higher verbal IQ seen among the students of interpretation. Additionally, the higher integrity in the anterior segment of the right arcuate fasciculus may indicate greater bilaterality of language processes among the Interpretation students as has been previously suggested (e.g., Darò, 1989; Gran & Fabbro, 1988). These group differences may suggest that greater white matter integrity in language related tracts allows for success in simultaneous interpretation.

These results, however, are in discord with the smaller FA values seen in professional interpreters (Elmer et al., 2011). While that study revealed differences primarily in the corpus callosum, the left insula, left frontal pole, and right inferior parietal lobule were also indicated. The left insula and frontal pole regions could belong to the uncinate fasciculus, while the inferior parietal lobule is an endpoint for both the anterior and posterior segments of the arcuate fasciculus. It is critical to note that the present study and the study by Elmer and colleagues used different methodologies. The present study made use of whole brain tractography and virtual dissections to consider individual tracts in their entirety. Thus, the measures examined applied to each tract specifically and wholly. The study by Elmer and colleagues, on the other hand, employed voxel-based analyses. While this method allows for specificity in the voxel location, it is ambiguous about the particular tracts causing the differences. In areas where multiple tracts converge, such as the insula-adjacent extreme capsule, it is impossible to separate the effects of individual tracts. Given these differences in methodology, speculation about the differing effects seen in the two studies is premature.

4.4.4 Effects of Training on Gray Matter Volume

Training in simultaneous interpretation appears to have a specific effect on regions which support the simultaneity of input and output. In a recent fMRI study examining the brain areas recruited during simultaneous interpretation, bilateral putamen and bilateral superior temporal gyrus were signaled as particularly involved in the simultaneity of speaking and listening during SI (Hervais-Adelman et al., 2014). This was evidenced as a modulation of activity with the duration of overlap of input and output in the simultaneous interpretation condition. Additionally, increased activity during speech shadowing compared to passive listening was seen, but no difference between the SI and shadowing conditions. Hervais-Adelman and colleagues posited that the putamen acts to suppress the speech production of the input language, in contrast to the caudate which functions to monitor and control language selection. Thus, the greater amassing of gray matter volume in bilateral putamen during training in SI in the present study may reflect the particular need during SI to inhibit the production of an input language. Indeed, the simultaneity of speech streams is a unique characteristic of simultaneous interpretation and therefore likely requires the development of new skills and structure supporting those skills.

As the superior temporal lobe is the locus of auditory cortex, the recruitment of these regions likely reflects the greater demands on auditory processing and attention during simultaneous interpretation. Activation was seen bilaterally in superior temporal lobe structures in the abovementioned study as well as in another fMRI study of simultaneous interpretation (Aherns et al., 2010). In the present study, however, the volumetric changes were left-lateralized. This difference may be due to the techniques used. FMRI studies reveal transient activation, whereas VBM analyses show areas of volumetric difference. Therefore, while activation may occur bilaterally, significant volumetric change may only occur in the language dominant left hemisphere. This seeming lateralization of auditory language processing may be reconciled with the previous suggestion of greater bilaterality of language by considering two points. First, the bilaterality may not apply equally to all aspects of language processing. Thus, while lexical retrieval and speech production may come to rely additionally on the right hemisphere, this may not pertain to auditory processing. Second, increased use of the right hemisphere in language, does not necessarily indicate equal use between the hemispheres, thus, lateralization may still be evident. In fact, changes in the right superior temporal lobe were present, but did not survive correction for multiple comparisons.

The present study evidenced additional changes in gray matter volume in the right thalamus and right calcarine sulcus. While these areas did not show a modulation of activity based on the degree of input-output overlap during SI in the study by Hervais-Adelman and colleagues, they did show greater activity during speech shadowing compared to passive listening in that same study. Thus, the volumetric changes seen in these regions may also be related to the simultaneity of speaking and listening. Alternatively, the right thalamus may be recruited when adjusting to multiple speakers compared to a single speaker. In an fMRI study of speech shadowing, a network including the right thalamus and temporal lobe regions was activated in conditions in which the input speech included multiple speakers (consecutively, not simultaneously), as opposed to a single speaker (Peschke, Ziegler, Kappes, & Baumgaertner, 2009). The authors theorized that these regions reflect a speaker normalization process. As the input in simultaneous interpretation comes from multiple speakers with varying accents, we could speculate that this region used for speaker normalization may have seen volumetric increases during SI training.

Another region showing a specific effect of training was the left temporal pole/fusiform gyrus area, which was an extension of the left putamen cluster. This region has previously been implicated in picture naming and verbal fluency (e.g., Abutalebi et al., 2014; Mion et al., 2010). Moreover, it is connected to the frontal pole through the uncinate fasciculus, which has been associated with lexical retrieval (Catani & Mesulam, 2008; Lu et al., 2002) and showed greater white matter integrity among the Interpretation students in the present study. Thus, the effect of training in the left temporal pole may be associated with greater demands on semantic processing and lexical retrieval in simultaneous interpretation.

A differential change between the Interpretation students and Translation students was also revealed in left superior occipital cortex. This region has not previously been implicated in simultaneous interpretation or its component processes; however, it has been connected to literacy and letter analysis (Carreiras et al., 2009; Fulbright, Manson, Skudlarski, Lacadie, & Gore, 2003). Thus, this training difference may be due to a pruning of the area among the students of translation. Similar to the smaller volumes among the students of interpretation in the spoken word learning regions, the students of translation may show improved efficiency specifically in brain areas which process written language.

Finally, the changes seen bilaterally in the putamen extended inferiorly to bilateral amygdala. The role of these areas in simultaneous interpretation is unclear. Additional studies are needed to replicate this result.

Overall, the changes in gray matter seen in connection with training in simultaneous interpretation suggest the growth of areas which support simultaneous language input and output. These results are in contrast to the smaller gray matter volumes in areas which support language control previously seen in professional interpreters compared to multilinguals (Elmer et al., 2014). These seemingly contrary results may be reconciled by assigning different roles to training and experience in simultaneous interpretation. As Elmer and colleagues suggested, professional experience with SI may lead to specialization and cortical pruning in the language control regions. Initial training in SI, on the other hand, may be defined by growth of brain regions that support previously untrained processes such as simultaneous input and output. Future longitudinal studies which follow interpreters through training and professional experience may help to clarify the roles of greater amassment and greater efficiency. It should be noted, however, that increased gray matter volume has typically been seen in connection with training as well as in professionals compared to amateurs across domains (e.g., Draganski et al., 2004; Gaser & Schlaug, 2003; Li et al., 2014). Thus, the finding of decreased volume among professional interpreters should be viewed with caution and replicated.

As a final note, the increased volume in regions supporting simultaneity of input and output language suggests that speech shadowing or other conditions which involve the simultaneity of input and output may not represent the optimal baseline for comparison to SI in functional neuroimaging studies of simultaneous interpretation. The use of these baselines is intended to isolate the SI process.

However, the fact that regions involved in simultaneity of speaking and listening, but not those involved in language control, are specifically affected by training in SI suggests the critical importance that managing input-output simultaneity has in SI.

4.4.5 Effects of Training on White Matter Integrity

Training in simultaneous interpretation was accompanied by a strengthening of the anterior and long segments of the left arcuate fasciculus. Both of these tracts provide a pathway between the left inferior frontal gyrus (IFG) and more posterior areas. It is notable that across the three studies which examined the brain areas activated during simultaneous interpretation, activation in the left IFG was consistently found (Aherns et al., 2010; Hervais-Adelman et al., 2014; Rinne et al., 2000). Though this area is known to be involved in language processes, the activation was seen in comparison to a shadowing or free speech condition suggesting a particular importance of left IFG during simultaneous interpretation. The increased integrity of the anterior and long segments in the left hemisphere may reflect an increased need for connectivity between left IFG and other areas during simultaneous interpretation. Interestingly, no change in gray matter volume was seen in the left IFG.

The strengthening of these left arcuate fasciculus segments in the students of interpretation was in addition to an already present advantage. In the right hemisphere, however, the advantage among the students of interpretation in the volume of the anterior segment of the arcuate fasciculus tapered during training with both groups moving toward the mean. The increase seen among the Translation students may be related to greater bilaterality of language processing, however, the cause for the decrease among the Interpretation students remains unclear.

Finally, a similar move toward the mean was seen in the fractional anisotropy of the right frontal aslant tract, though the initial difference was not significant. This tract also evidenced increased volume between Phase 1 and Phase 2 across the groups. Thus, the differences in FA values may relate to the specific nature of the volumetric increase.

4.4.6 Limitations and Future Directions

The present study represents the first examination of the structural changes associated with training in simultaneous interpretation. While it adds to the literature on neuroplasticity in relation to skill learning, it does have some limitations which should be addressed in future studies. Similar to the study presented in the previous chapter, the sample size was small. Additionally, the number of participants in each group was unbalanced. Both of these issues could be addressed by collaborating with training programs as suggested in the previous chapter. Further, participants were scanned in the first three months of training, rather than prior to any training. As mentioned in the previous chapter, this could potentially lead to the obscuring of differences due to training, but it is unlikely that it would cause spurious differences to appear. Future studies, however, should aim to acquire a more accurate baseline.

The results of the present study together with those of previous studies on professional interpreters suggest divergent roles for initial SI training and professional SI experience on brain structure. Thus, future studies may wish to examine aspiring interpreters through training and into professional experience to better understand the role each experience plays in sculpting the brain.

The choice of students of translation as the comparison group in the present study controlled for the potentially confounding factors of language proficiency and intense language training. These students, however, were involved in their own training program which may lead to specific, but different changes in brain structure. The addition of another control group composed of either multilinguals or monolinguals may aid in differentiating the effects of training in interpretation versus training in translation.

Finally, in the present study we examined five white matter tracts previously associated with spoken language. However, given the complex nature of simultaneous interpretation, changes may be seen in other tracts. In particular, examination of the corpus callosum may reveal interesting differences as it was indicated in the study on professional interpreters (Elmer et al., 2011) and may speak to questions of symmetric versus asymmetric language processing. Additionally, our understanding of the circuits used in simultaneous interpretation would be furthered by the assessing functional connectivity in addition to structural connectivity.

4.5 Concluding Remarks

The present study indicates that initial training in simultaneous interpretation is associated with physical changes in brain regions which support the component processes of this skill. In particular, gray matter areas thought to support simultaneous language input and output saw growth and white matter pathways in the left hemisphere which underlie language were strengthened. Further, an efficient spoken word learning network and strong language pathways may be linked to an initial aptitude for simultaneous interpretation. Examinations of the structural changes which accompany training in simultaneous interpretation are highly informative, not only about the specific processes used when interpreting, but also about skill learning more broadly.

CHAPTER 5

TO INHIBIT OR NOT TO INHIBIT?

5.1 Introduction

The nature of language management is of critical importance in the bilingualism field as the constant management of two languages is typically credited as the source of the cognitive benefits seen in bilinguals. An early theory, which is still commonly cited today, suggested that language control relies on the inhibition of the unintended language through the selection of language task schemas (Green, 1998). An alternative model posited that bottom-up activation and mutual inhibition underlies language control (Dijkstra & van Heuven, 1998). More recently, however, Green and Abutalebi (2013) proposed that language control processes are not uniform across bilinguals, but rather are adaptive to the common interactional contexts used by each bilingual. In detailing three such interactional contexts, single-language, dual-language, and dense code-switching, they noted that the languages are in competition in the first two situations, while in dense code-switching the languages exist cooperatively.

To further decompose the language control processes, Green and Abutalebi (2013) identified eight processes which may be more or less engaged based on the interactional context. Among these processes is interference control, which is supported largely by inhibitory processes. Large demands are placed on interference control in contexts in which the two languages are in competition, however, the demands are neutral when the languages exist cooperatively. Thus, cooperative language usage may be marked by less inhibition of the languages.

Though the simultaneous interpretation context was not detailed by Green and Abutalebi (2013), we may conjecture that a cooperative relationship between the languages exists in this context as well. A similar explanation of language management in simultaneous interpretation was offered by Ibáñez and colleagues (2010). These authors suggested that the languages of an interpreter are not in competition, but instead that both are maintained active. This supposition is further supported by the enhanced sustained control of professional interpreters demonstrated in Chapter 2 (Section 2.3.4.2). As reduced inhibition may signal a cooperative language context, directly assessing the inhibition applied to recently abandoned languages in interpreters may provide further support for a cooperative, rather than competitive, language context.

Inhibition is typically thought of as a singular construct; however, multiple types of inhibition have been noted (Nigg, 2000) and dissociated. Friedman and Miyake (2004) examined three types of inhibition: prepotent response inhibition, resistance to distractor interference, and resistance to proactive interference. The first two types of inhibition are associated with the ANT and Stroop tasks, while resistance to proactive interference is more closely associated with inhibiting an abandoned task set (or language). Friedman and Miyake (2004) found that resistance to proactive interference was statistically separable from prepotent response inhibition and resistance to distractor interference, which were represented by a single latent variable. Thus, the lack of differences in inhibitory control between interpreters and multilinguals seen in Chapter 2 and before and after training in Chapter 3 (tested with ANT and Stroop tasks) does not directly speak to the current supposition.

To examine the inhibition used when abandoning a language we employed the n-2 repetition cost. This measure was first introduced by Mayr and Keele (2000) in an investigation of the inhibition applied to previous task sets during intentional shifts. The premise of the measure is that returning to a recently inhibited task should be more difficult and therefore cause a decrement in performance. To quantify this decrement the authors classified trials in a three task switching paradigm based on the task presented two trials previously (the *n-2* trial). Thus, on "n-2 repetition" trials participants performed the same task on trial *n* and trial *n-2* (e.g., Task 1 – Task 2 – Task 1), while on "n-2 non-repetition" trials the tasks used on trials *n* and *n-2* differed (e.g., Task 3 – Task 2 – Task 1). The finding that responses on n-2 repetition trials are slower and more error prone has been widely replicated (e.g., Arbuthnott & Frank, 2000; Arbuthnott, 2008; Gade & Koch, 2012; Houghton, Pritchard, & Grange, 2009; Philipp & Koch, 2006; Schuch & Koch, 2003). Further, the n-2 repetition cost (the difference in performance between n-2 repetition and non-repetition trials) has been recognized as the empirical signature of inhibitory processes in task switching paradigms (Kiesel et al., 2010; Koch, Gade, Schuch, & Philipp, 2010).

Despite the strength of the n-2 repetition cost in pinpointing inhibitory control processes, which are thought to be enhanced by bilingualism, only one study has examined the influence of bilingualism on n-2 repetition costs in the non-linguistic domain. Prior (2012) found larger n-2 repetition costs for bilinguals than for monolinguals and suggested that bilinguals apply more inhibition to recently abandoned tasks. These results, in conjunction with previous studies showing a bilingual advantage in inhibitory control (e.g., Bialystok et al., 2008; Costa et al., 2008), further underscore the dissociation of types of inhibition.

In the language domain, four studies have examined n-2 repetition costs using language switching paradigms; naturally comparisons to monolinguals are not feasible in these cases. The work by Philipp, Gade, and Koch (2007) was the first of these studies and confirmed the validity of the n-2 repetition cost measure in the language domain. The authors were additionally interested in the effect of language dominance on the size of the n-2 repetition cost, in the vein of findings of larger switching costs for dominant languages (e.g., Meuter & Allport, 1999). The results demonstrated a reliable n-2 repetition cost when switching among languages, confirming the use of inhibitory mechanisms in language switching. The effects of language dominance on n-2 repetition cost, however, were ambiguous. The cost was largest in L1, however, L3 showed a larger cost than L2. The influence of language dominance on n-2 repetition costs is further complicated by the findings of Guo, Liu, Chen, and Li (2013). Across three experimental manipulations these authors consistently found a larger cost in L2 than in L3. Additionally, Philipp and Koch (2009) found no interaction between language and trial type in two experiments, though both experiments included additional manipulations which may have obscured these interactions in the statistical analyses. Visual inspection of the data in the conditions most similar to those used in the current study suggests smaller n-2 repetition costs in L1. Finally, Branzi (2014) compared language and non-language switching tasks within a multilingual sample and found no correlation of n-2 repetition cost across the two task versions. That study, however, did not examine the effects of language within the language switching task. Given the diversity of results regarding n-2 repetition cost size, it appears that language dominance is not related to the cost in a straightforward manner. Instead, the size of the n-2 repetition cost may be affected by other language characteristics.

One potential influence on n-2 repetition costs is the lexical robustness of each of the languages. Lexical robustness refers to the familiarity with lexical representations and their frequency of access which leads to greater automaticity of retrieval (Costa, Santesteban, & Ivanova, 2006). In a study addressing trilingual speech production, Schwieter and Sunderman (2011) found relationships between lexical robustness and response times and switching costs in a language switching paradigm. In particular, larger L2 lexical robustness was associated with faster responses across languages, while larger L3 lexical robustness was associated with faster responses and larger switching costs in the L3. Though that study used a trilingual language switching paradigm, the authors did not consider n-2 repetition costs.

The present study investigated the link between the simultaneous use of two languages needed for simultaneous interpretation and inhibition of abandoned task sets and languages. To this end we examined students earning a Master in Conference Interpreting using a cross-sectional design with the aim of understanding how training in simultaneous interpretation affects inhibitory control. In the main experiment we targeted the training period by examining students at the beginning and end of a Master in Conference Interpreting, using the same time points tested in the longitudinal study presented in Chapter 3. This design, as opposed to comparing to age and education matched multilinguals, controls for the potential confounds related to selection into a training program in interpretation. For example, in Chapter 3 (Section 3.2.2.2), we showed that students selected for a training program in simultaneous interpretation have greater verbal intelligence. While this known difference could be taken into consideration during participant selection, additional untested differences may exist. The current sample selection limits the effects of these innate differences as all participants chose to study interpretation and were admitted to a training program. In the follow-up experiment students completing the coursework for the Master's program or a Triennale (equivalent to a Bachelor's degree) in Languages were added to the sample.

The study included a trilingual language switching paradigm and a three task switching paradigm. The use of language and non-language versions allowed us to examine differences in inhibitory control in the language domain as well as look for domain-general changes. These changes may be expected given the advantage interpreters displayed in mixing cost and overall response time on a non-linguistic task-switching paradigm (see Section 2.3.4.2). The two tasks were designed to be as comparable as possible and to maximize the sensitivity of the n-2 repetition cost measure. Additionally, to assess the influence of lexical robustness on the n-2 repetition cost, a verbal fluency task was included. As mentioned in the General Introduction (Section 1.2.1), professional interpreters have been previously shown to have an advantage in lexical robustness (Stavrakaki et al., 2012). Thus, the inclusion of the verbal fluency task had the further benefit of allowing the examination of the source of these differences, that is, whether they are innate or due to training.

We theorized that if interpreters maintain two languages by using less inhibition, and that this capacity is developed through SI training, then the students who had completed a Master in Conference Interpreting would exhibit a smaller n-2 repetition cost in the language switching paradigm with respect to a control group of students who were about to start this training. Further, if this modulation of inhibitory control acts domain-generally, the same effect would be expected on the non-language switching paradigm.

5.2 Main Experiment: Methods

5.2.1 Participants

The participants were students who were either starting or finishing a two-year Master's program in Conference Interpreting at the University of Trieste. Twenty-one students at the beginning of the Master's program participated (19 females). At the time of testing these students had attended one to two months of courses in the program. Twenty-two students finishing the Master's program were also recruited (14 females). These students had finished the coursework required for the program and most were working on their theses at the time of testing. All participants had no known neurological or psychiatric problems and reported normal color vision, which was confirmed with the Ishihara Color Vision Test (Ishihara, 1972). As biographical factors can influence measures of cognitive control, the two groups were matched on intelligence (measured with Raven's Advanced Progressive Matrices Series I; Raven, Raven, & Court, 1998) and socioeconomic status (measured with mother's years of education, Gottfried, Gottfried, Bathurst, Guerin, & Parramore, 2003; Noble, McCandliss, & Farah, 2007; Prior & Gollan, 2011; Stevens, Lauinger, & Neville, 2009; Table 14). Given the cross-sectional design of the study, it was not possible to match the groups on age and years of education. All participants gave written informed consent and were compensated for their time. The study was approved by the ethical committee of the Scuola Internazionale Superiore di Studi Avanzati (SISSA).

5.2.2 Tasks and Procedure

Participants were tested individually in a sound-attenuated booth during a single session. During the session participants completed a language history questionnaire, Series I of Raven's Advanced Progressive Matrices, a verbal fluency task, and the two switching tasks (language and non-language versions presented in a counterbalanced order). The verbal fluency and language switching tasks required the selection of an L1, L2, and L3 for each participant. Italian was always assigned as L1 and was a native language for all but one participant. The L2 was the language that each participant felt most confident in after Italian and the L3 the most comfortable language after that (languages used: Dutch, English, French, German, Russian, and Spanish). In cases where a participant was equally comfortable in two non-native languages, selections were made to avoid phonologically similar languages (i.e., Italian/Spanish and Dutch/German) and based on the participant's most recent exam scores.

5.2.2.1 Language history questionnaire. Participants were asked to provide information about all of the languages they knew and/or studied. For each language they were asked to detail how and when they learned the language, including immersion experiences, as well as to provide a self-rating in the areas of reading, writing, speaking, and understanding on a 5-point Likert scale. Additionally, the participants completed a questionnaire developed to identify functional fluency for all their non-native languages. Functional fluency was operationalized as a B2 level or above in the Common European Framework of Reference for Languages (CEF). The questionnaire asked participants to give their CEF level and respond to eight yes-or-no questions which targeted the B1-B2 border (see Appendix A for questionnaire items). The questionnaire contained two items for each of the four abilities (reading, writing, speaking, and oral comprehension), one item focused on academic usage and the other on personal usage. Participants were considered functionally fluent in languages for which they responded yes to seven or eight items. Finally, participants were also asked to evaluate how often they switched between languages within a conversation in the two years prior to testing using a 5-point Likert scale. This question was posed for the following seven situations: at home, with friends, at school/work, thinking, dreaming, talking to oneself, and expressing anger and affection.

Table 14: Biographical and language characteristics of participants in Main Experiment

	Beginning Master's (N=21)	Writing Thesis (N=22)	*p*-value
Age (in years)	22.7 (1.8)	24.6 (1.1)	$p < .001$
Years of education	16.3 (0.6)	18.3 (0.7)	$p < .001$
Raven's APM score	10.7 (1.5)	11.0 (1.4)	$p = .462$
Mother's years of education	13.7 (2.8)	13.5 (3.4)	$p = .825$
Number of functional languages	3.5 (0.5)	3.7 (0.7)	$p = .334$
L2 reading	4.3 (0.5)	4.8 (0.4)	$p = .002$
L2 writing	3.9 (0.4)	4.2 (0.6)	$p = .054$
L2 speaking	4.2 (0.5)	4.4 (0.7)	$p = .216$
L2 understanding	4.3 (0.4)	4.7 (0.5)	$p = .002$
L3 reading	4.1 (0.7)	4.4 (0.6)	$p = .105$
L3 writing	3.5 (0.7)	3.9 (0.8)	$p = .102$
L3 speaking	3.5 (0.7)	3.8 (0.7)	$p = .270$
L3 understanding	3.9 (0.5)	4.4 (0.6)	$p = .009$

Note: Values reported are means with standard deviations in parentheses. For one participant no language data was available and for another participant L2 self-ratings were missing.

As expected given the additional two years of language instruction, the Thesis group reported higher self-ratings than the Beginning Master's group for many of the areas in both L2 and L3 (see Table 14). However, the two groups did not differ on their self-ratings in either L2 or L3 in the area of speaking, which was the area most strongly utilized in the completed tasks. Further, across the groups L2 self-ratings were higher than L3 ratings for all four areas (*p*s < .001). Additionally, the two groups did not differ in their number of functional languages. Finally, the groups did not differ in code-switching frequency in all situations (*p*s ≥ .259) except thinking, where the Beginning Master's students code-switched more frequently (*p* = .015).

5.2.2.2 Raven's Advanced Progressive Matrices. Non-verbal intelligence was measured using Series I of Raven's Advanced Progressive Matrices (Raven et al., 1998). In this task, participants view patterns in a 3x3 matrix, each of which is missing a piece, and must choose the piece that completes the pattern from eight options. Series I contains 12 items and participants completed the task untimed, though most took 5-10 minutes. In an effort to limit the total time of the testing session, we opted to only use Series I, which has been previously used to provide a quick measure of intelligence (e.g., Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Paap & Greenberg, 2013).

5.2.2.3 Verbal fluency task. Verbal fluency was measured in the L1, L2, and L3 for each participant. Each language was tested in a separate block and the order of the languages was randomly selected. Three semantic categories and three letter categories were presented for each language. The nine semantic categories used throughout the task were grouped into three sets of three categories (adapted from Schwieter & Sunderman, 2011; see Appendix B for all categories used). The same was true of the nine letter categories. For each participant a set of semantic categories and a set of letter categories were randomly assigned to each language, with no repetition of categories across languages. Each block began by announcing the language to use in the following categories (language names were written in Italian). Each category began with fixation cross for 500 ms, followed by a 100 ms beep (400 Hz) and then the category name written in the language to be used in the block.

Participants were given 60 s to name as many items belonging to the category as possible. The conclusion of the 60 s was signaled with another 100 ms beep. A blank screen appeared for 3000 ms before the next category. The six categories presented in each block appeared in a random order. A break occurred at the end of each block and participants could initiate the next block through a button press. Participants were instructed that they would see two types of categories (semantic and letter) and have 60 seconds to name as many items as they could in the category. They were also told that repetitions, words with the same root, and proper names would not be counted. Before completing the experimental categories, participants completed two practice categories (P and kitchen items-*cose nella cucina*) in Italian with an experimenter present to clarify any difficulties.

The entire task was recorded using a digital recorder for offline transcription and coding. Transcription and coding were completed by a native speaker of each language. All responses initiated within the 60 second time limit for each category were transcribed and marked as either countable items or items to be excluded. For both category types responses were excluded that were: not a word in the language, not part of the category, repetitions of a previous response in the same category, and proper names. Further for semantic categories repetitions of a concept (e.g., eggplant and aubergine) and superordinates when subordinates were also named (e.g., bird, pigeon, bluejay) were excluded. Finally, for letter categories, only one word with a given root was counted (e.g., fast, faster, fastest).

5.2.2.4 Language switching task. Participants viewed a series of stimuli composed of the letter X, the # sign, and a digit between 2 and 9 (the digit 1 was excluded due to high phonological similarity across the languages used). The X and # were not informative, but rather were included to match the visual complexity of the stimuli between the language and non-language tasks. The stimuli components were black and were presented with equal probability in each of the six possible orders (e.g., X#2, 2#X). Participants were asked to name aloud the digit in their L1, L2, or L3 according to the cue presented. Cues were black frames surrounding the stimulus in the shape of a diamond, a hexagon, and a triangle (see Figure 11 for example item). Graphic cues were chosen because they have been previously associated with larger n-2 repetition costs (Guo et al., 2013; Houghton et al., 2009). The cue-language pairings were counterbalanced across participants. A visual reminder of these pairings was visible throughout the experiment. Each stimulus was categorized as either an n-2 repetition or non-repetition trial; the difference between these trial types quantifies inhibition. On n-2 repetition trials, the language used on the current trial was the same as that used on the n-2 trial (e.g., English – Italian – English). Thus participants were returning to a recently inhibited language on these trials. Conversely, on n-2 non-repetition trials, the current language differed from that used on the n-2 trial (e.g., French – Italian – English). Immediate language repetitions were excluded from the task design since their presence has been associated with a decrease in n-2 repetition cost (Philipp & Koch, 2006). Consequently, n-2 non-repetition trials made use of all three languages and n-2 repetition trials of two languages.

Each trial began with a 500 ms blank screen followed by cue presentation. The target stimulus appeared inside the cue 100 ms later. A short cue-to-stimulus interval was employed to enhance n-2 repetition costs (Guo et al., 2013; Philipp et al., 2007). The stimulus and cue remained onscreen until 200 ms after the participants started their vocal response, at which point the trial ended. The onset of vocal responses was recorded with a voice key. Before the task instructions were given a microphone sensitivity test was completed in which the participant pronounced the numbers 2 to 9 in each of their three languages. The microphone was adjusted and the test repeated until all possible responses successfully triggered the voice key. The experiment began with a practice session consisting of 30 trials, which could be repeated until the participant felt comfortable with the task paradigm (two participants, one from each group, repeated the practice once). The experimental session consisted of six blocks, each with 120 trials. After every block there was a break and participants could initiate the next block with a button press. The sequence of trials was pseudorandomized for each participant according to the following restrictions. There were an equal number of n-2 repetition and non-repetition trials, as well as an equal number of trials for each language, divided approximately evenly between n-2 repetition and non-repetition trials. Further, in each block of 120 trials, each digit was presented 15 times, 5 times in each language. Immediate language repetitions and immediate digit repetitions were excluded. Finally, a digit could not appear with the current language if it had been used on the most recent trial using that language.

Participant responses were recorded using a digital recorder. The recording was used offline to code the accuracy of responses. To aid in the alignment of responses and trial number during offline coding, a 100 ms beep during cue presentation, unheard by participants, was fed directly to the digital recorder. Responses were coded for the accuracy of the response (i.e., the correct number produced in the correct language) and for false starts, when a sound other than a response triggered the voice key (e.g., a cough, a speech filler, a corrected response from the previous trial).

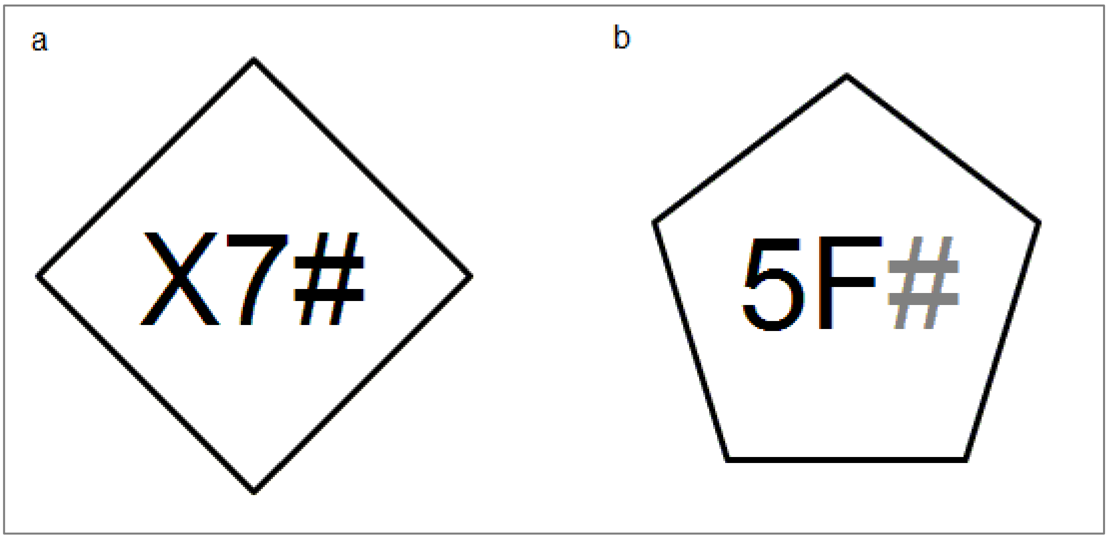

Figure 11: Example items for (a) the language switching task and (b) the non-language switching task.

5.2.2.5 Non-language switching task. The stimuli in this task were composed of a black letter (A-H), a black digit (2 to 9), and a colored # sign (black-*nero*, gray-*grigio*, blue-*blu*, green-*verde*, red-*rosso*, yellow-*giallo*, pink-*rosa*, brown-*marrone*). To allow for comparability in the stimulus set size between this and the language switching task, eight composite stimuli were created for each participant. The composite stimuli grouped a letter, a number, and a color which appeared together throughout the experiment with equal probability in each of the six possible orders (e.g., letter-number-#, number-#-letter). Participants were asked to name aloud in Italian the letter, the digit, or the color of the # sign according to the cue presented. Cues were black frames surrounding the stimulus in the shape of a square, a circle, and a pentagon (see Figure 11 for example item). The cue-task pairings were counterbalanced across participants. A visual reminder of these pairings was visible throughout the experiment. As in the language switching task, each stimulus was classified as either an n-2 repetition trial (e.g., number – color – number) or an n-2 non-repetition trial (e.g., letter – color – number). The details of the task presentation, pseudorandomization, and offline coding were identical to those used in the language switching task. During the microphone sensitivity test, participants named the eight letters, digits, and colors presented in the task. This test was additionally used to check that the participants correctly named the colors. As with the language switching task, participants could repeat the practice session until they felt confident in the task (two Beginning Master's participants repeated the practice once).

5.3 Main Experiment: Results

5.3.1 Verbal Fluency

Data from five participants were excluded from analyses on this task. Four participants (one Beginning Master's student and three Thesis students) had grown up speaking two languages and one participant (Beginning Master's student) was not a native Italian speaker. These participants were excluded because the following analyses make distinctions between L1, L2, and L3 and these classifications do not match the language experiences and representations of these participants.

To analyze differences in verbal fluency a mixed effects three-way ANOVA with category type (semantic, letter) and language (L1, L2, L3) as within-subjects factors and group (Beginning Master's students, Thesis students) as a between-subjects factor was used. Reported results reflect a Greenhouse-Geisser correction.

Participants produced more words in response to semantic categories than letter categories ($F(1,36) = 19.560$, $p < .001$, $\eta_p^2 = .352$; Figure 12). The number of responses was also affected by language ($F(1.867,67.202) = 126.838$, $p < .001$, $\eta_p^2 = .779$) with the most responses produced in L1 followed by L2 and then L3 (ps < .001 from post hoc t-tests evaluated at $\alpha = .017$ to correct for multiple comparisons). There was no overall difference between the groups ($p = .987$). Though the language and group factors showed a marginal interaction ($F(1.867,67.202) = 3.110$, $p = .054$, $\eta_p^2 = .080$), post hoc t-tests revealed no significant group differences (ps ≥ .199). No other interactions were significant (ps ≥ .172).

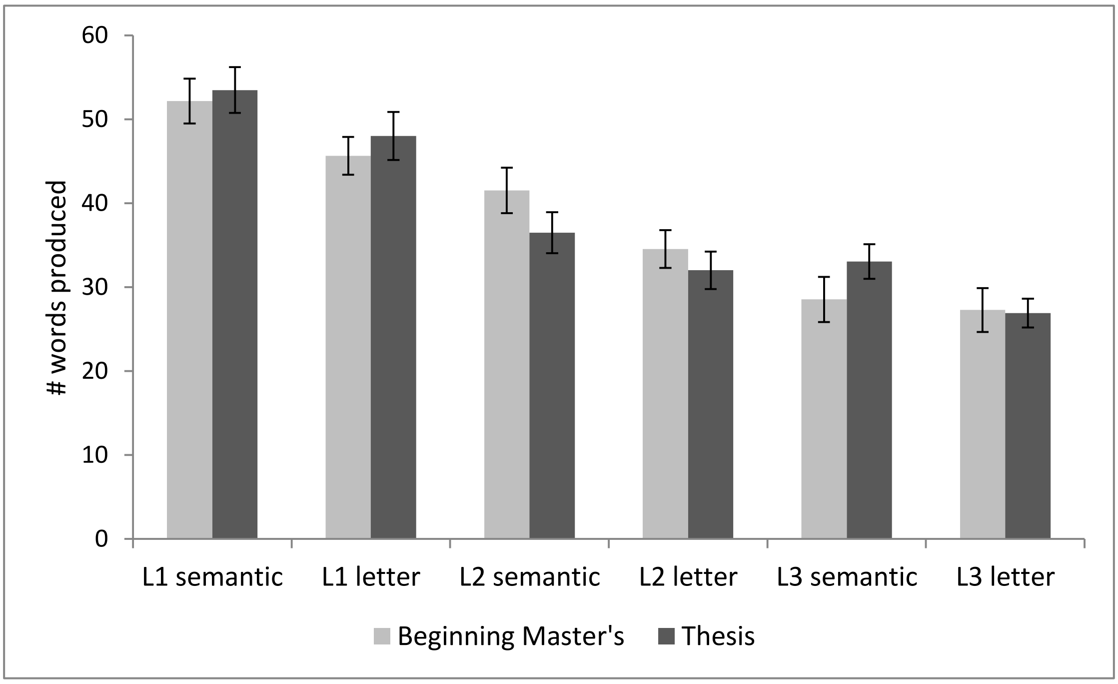

Figure 12: Verbal fluency results for language and category type by group in the Main Experiment. Error bars represent the standard errors of the mean.

5.3.2 Language Switching Task

Data from six participants were excluded from all analyses on this task. The five participants (two Beginning Master's students and three Thesis students) who spoke languages other than Italian (or Italian dialects) from birth were excluded because the following analyses also make distinctions between L1, L2, and L3. A final participant (Thesis student) was excluded due to a disruption during the task. The exclusion of these participants did not change the results of the comparisons between the two groups on the biographical and language characteristics mentioned above.

Non-parametric tests were used to analyze the accuracy data given that these data were not normally distributed. The Mann-Whitney U test was used to compare the two groups and the Wilcoxon signed-ranked test was used to compare conditions. Response time (RT) data were analyzed using a mixed effects three-way ANOVA with trial type (n-2 repetition, n-2 non-repetition) and language of the current trial (L1, L2, L3) as within-subjects factors and group (Beginning Master's students, Thesis students) as a between-subjects factor. For these analyses error trials and the two trials following an error, to ensure the correct trial type assignment, were excluded. Trials with a false start were also excluded since RT was not an accurate reflection of performance in these trials. Additionally, for each participant, trials with an RT more than 3 standard deviations from their individual mean were excluded. Finally, responses that were faster than 200 ms and slower than 3500 ms were excluded. This trimming procedure resulted in the exclusion of 12.6% of all trials. Reported results reflect a Greenhouse-Geisser correction.

Analyses on accuracy revealed no overall difference between the groups (p = .298). Further, there were no differences between the groups when considering each language and trial type separately (ps ≥ .271). Across the two groups, accuracy was higher on n-2 non-repetition trials compared to n-2 repetition trials (Wilcoxon T = 417.5, Z = 2.055, p = .040). Additionally, accuracy was higher when responding in L1 than in L2 and L3 (p < .001 and p = .003, respectively), with no difference between L2 and L3 (p = .976).

The RT analysis showed a main effect of trial type (F(1,35) = 45.117, p < .001, η_p^2 = .563, Figure 13) with faster responses to n-2 non-repetition trials than n-2 repetition trials. The main effect of language was also significant (F(1.819,63.656) = 14.053, p < .001, η_p^2 = .286, Figure 13). Post hoc t-tests (evaluated at α = .017 to correct for multiple comparisons) revealed that responses in L1 were faster than in L2 and L3 (t(36) = 3.827, p < .001 and t(36) = 5.540, p < .001, respectively), with no difference between L2 and L3 (p = .589). These two variables, trial type and language, also showed a significant interaction (F(1.990,69.649) = 4.075, p = .021, η_p^2 = .104). There was no main effect of group (p = .920) or significant interactions between group and either the trial type or language variables (p = .777 and p = .621, respectively). However, the 3-way interaction of trial type, language, and group was significant (F(1.990,69.649) = 5.701, p = .005, η_p^2 = .140, see Figure 14).

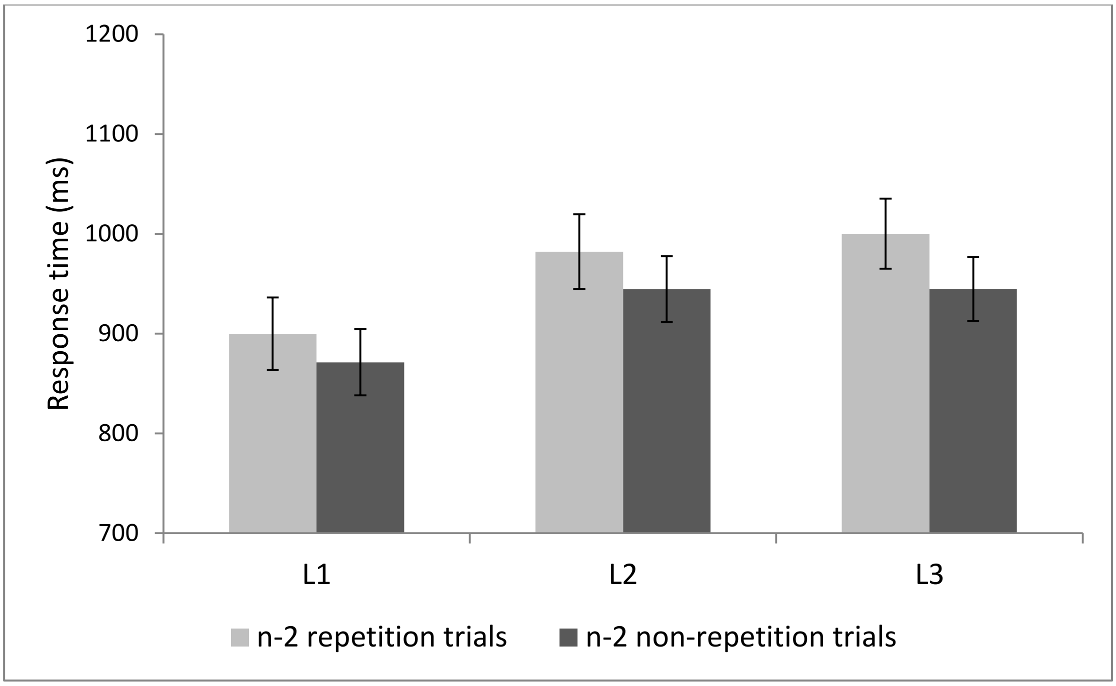

Figure 13: Response times for each language by trial type on the language switching task in the Main Experiment. Error bars represent the standard errors of the mean.

Post hoc *t*-tests (evaluated at $\alpha = .0083$ to correct for multiple comparisons) comparing the n-2 repetition cost between languages within each group were used to decompose the 3-way interaction. These comparisons revealed that within the Beginning Master's group the n-2 repetition cost in L1 was smaller than in L2 and L3 ($t(18) = 3.312$, $p = .004$ and $t(18) = 3.530$, $p = .002$, respectively), which did not differ from one another ($p = .761$). Within the Thesis group, however, there were no differences in n-2 repetition cost between the languages (ps $\geq .060$). When post hoc *t*-tests are considered as comparisons of the n-2 repetition cost between the groups within each language (evaluated at $\alpha = .017$ to correct for multiple comparisons), no difference was seen in L2 and L3 ($p = .129$ and $p = .918$, respectively), however a trend for a smaller n-2 repetition cost for Beginning Master's students was seen in the L1 ($t(35) = 2.226$, $p = .033$).

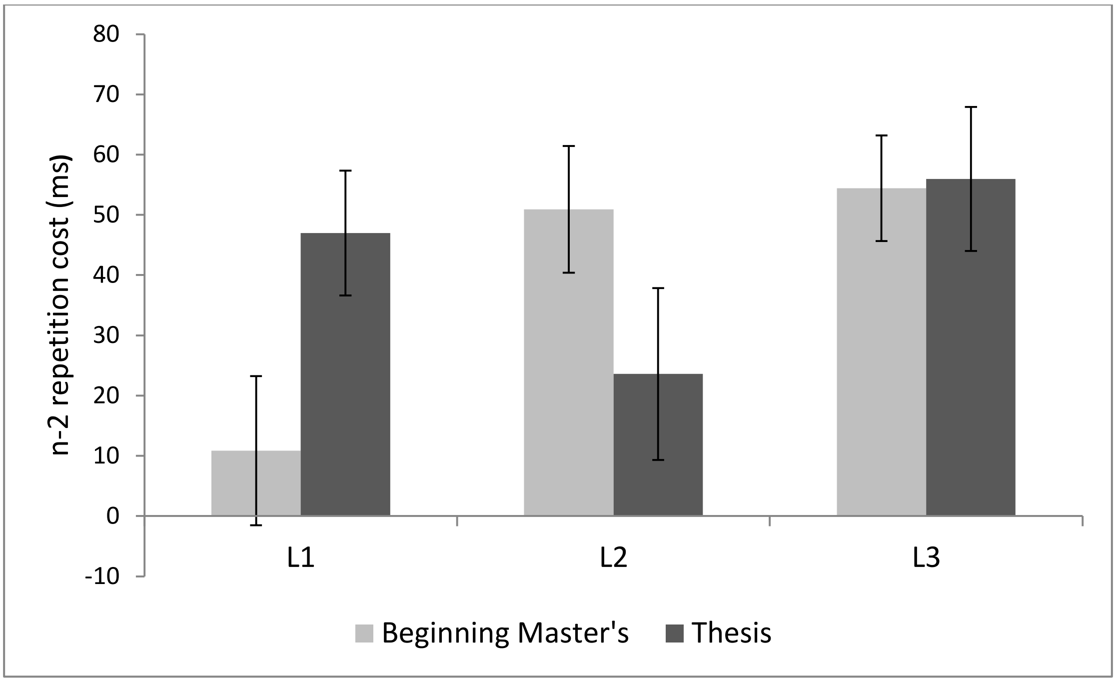

Figure 14: n-2 repetition costs for each language by group on the language switching task in the Main Experiment. Error bars represent the standard errors of the mean.

To investigate the role of language characteristics on n-2 repetition costs correlational analyses were conducted. The n-2 repetition cost in L1, L2, and L3 were tested for correlations with the semantic and letter fluency scores in L1, L2, and L3 as well as the self-ratings in L2 and L3. There were no significant correlations between n-2 repetition cost and any of the verbal fluency or self-rating measures (*p*s $\geq .185$; these values were not corrected for multiple comparisons as these analyses were exploratory in nature).

5.3.3 Non-Language Switching Task

Data from seven participants were excluded from all analyses on this task. The five participants (two Beginning Master’s students and three Thesis students) who spoke languages other than Italian (or Italian dialects) from birth were excluded as early bilingualism may affect n-2 repetition costs (Prior, 2012). One additional participant (Beginning Master's student) was excluded due to difficulties coding her data offline. A final participant (Beginning Master's student) was identified as an extreme outlier within her group based on her accuracy rate (more than 3 interquartile ranges below the 1st quartile) and therefore excluded. The exclusion of these participants did not change the results of the comparisons between the two groups on the biographical and language characteristics mentioned above.

Non-parametric tests were used to analyze the accuracy data given that these data were not normally distributed. The Mann-Whitney U test was used to compare the two groups and the Wilcoxon signed-ranked test was used to compare conditions. Response time data were analyzed using a mixed effects three-way ANOVA with trial type (n-2 repetition, n-2 non-repetition) and task of the current trial (color, letter, number) as within-subjects factors and group (Beginning Master's students, Thesis students) as a between-subjects factor. For these analyses error trials and the two trials following an error, to ensure the correct trial type assignment, were excluded. Trials with a false start were also excluded since RT was not an accurate reflection of performance in these trials. Additionally, for each participant, trials with an RT more than 3 standard deviations from their individual mean were excluded. Finally, responses that were faster than 200 ms and slower than 3500 ms were excluded. This trimming procedure resulted in the exclusion of 10.3% of all trials. Reported results reflect a Greenhouse-Geisser correction.

Analyses on accuracy revealed no overall difference between the groups ($p = .146$). Further, the groups did not differ on n-2 repetition trials, n-2 non-repetition trials, letter trials, or number trials (ps $\geq .121$). However, the Beginning Master's group were more accurate on color trials (Mann-Whitney U = 241.5, z = 2.546,

p = .010). Across the two groups, accuracy was higher on n-2 non-repetition trials compared to n-2 repetition trials (Wilcoxon T = 404, Z = 2.628, p = .009). There were no differences in accuracy between the three tasks (ps ≥ .220).

The RT analysis showed a main effect of trial type ($F(1,34) = 224.283$, $p < .001$, $\eta_p^2 = .868$, Figure 15) with faster responses to n-2 non-repetition trials than n-2 repetition trials. No other effects or interactions were significant (ps ≥ .157).

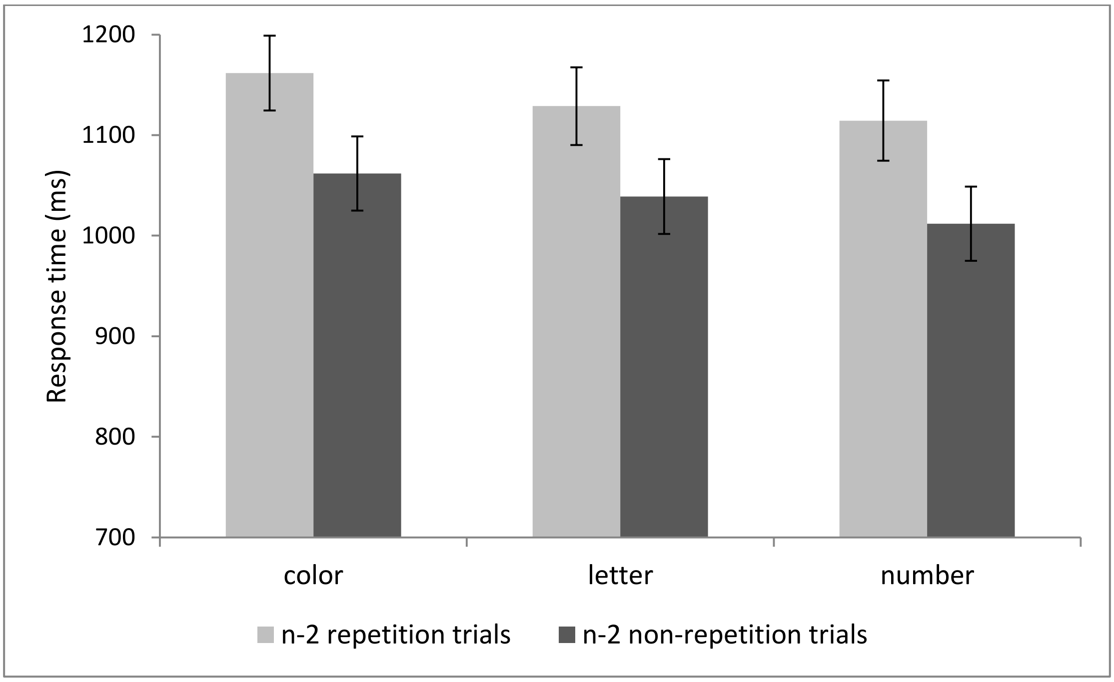

Figure 15: Response times for each task by trial type on the non-language switching task in the Main Experiment. Error bars represent the standard errors of the mean.

5.4 Main Experiment: Discussion

5.4.1 Summary of Results

This experiment investigated the effect of training in simultaneous interpretation on lexical robustness and inhibitory control in the language and non-language domains using a cross-sectional design. Students at the beginning and the end of a Master in Conference Interpreting were compared. Lexical robustness, measured with a verbal fluency task, did not differ between the two groups and was shown to be largest in the L1 and smallest in the L3. Inhibitory control was measured using the n-2 repetition cost in a trilingual language switching paradigm and a three task switching paradigm. In the language task the amount of inhibitory control employed was affected by the combination of group and language used. The students at the beginning of the Master's showed smaller n-2 repetition costs in L1 than L2 and L3, while the students at the end of the Master's showed no differences between the languages. The two groups did not differ on the non-language task.

5.4.2 Lexical Robustness

As mentioned in the General Introduction (Section 1.2.1) and Introduction of the current chapter, an advantage for professional interpreters in verbal fluency has been previously noted (Stavrakaki et al., 2012). The origin of this advantage, however, remains unclear; it may be an innate characteristic that allowed success or a characteristic that was acquired through training. The results of the current experiment, which showed no difference in verbal fluency between students beginning a Master of Conference Interpreting and those finishing the same Master's program, suggest that verbal fluency is not enhanced through training in simultaneous interpretation. This ability, instead, may fall into the category of innate characteristics that allow success.

5.4.3 Inhibitory Control

The present results replicated the finding of a reliable n-2 repetition cost in language switching, confirming the use of inhibitory control mechanisms when switching between languages. The results further indicated that the two groups differed in the amount of inhibition used depending on the language. The Beginning Master's students inhibited the L1 less than the Thesis students did, while the opposite pattern was observed (though not significant) in the L2. The differences in n-2 repetition cost were not explained by either lexical robustness in the three languages or self-rated ability in the L2 and L3.

A smaller n-2 repetition cost for the untrained group, as seen in the L1, runs contrary to the hypothesis that simultaneous use of two languages is managed through less inhibition. However, it is noteworthy that the Beginning Master's group had a very small n-2 repetition cost in L1 (11 ms) when compared with the values reported by previous studies in comparable conditions (106 ms in Philipp et al., 2007 and 89 ms in Guo et al., 2013). Therefore, it is unclear whether the Beginning Master's students represent a true baseline from which the Thesis students have shifted. Indeed, the Beginning Master's students had attended one to two months of courses at the time of testing, including courses on simultaneous interpretation. In contrast, the Thesis students were focused on their theses at the time of testing, and not necessarily actively practicing simultaneous interpretation. In light of these facts, the reduced n-2 repetition cost may be related to recent practice with SI.

Two accounts of the present data are available. In the first, the Thesis group has changed their level of inhibition after training with SI, leading to a larger inhibition of the L1, contrary to the original prediction. The higher level of inhibition specifically in the L1 may reflect a particular need to suppress the dominant L1 when interpreting into a non-dominant language. The second account suggests that training with SI does lead to a decrease in inhibition, at least in the L1, but the effect is related to recent practice and dissipates during periods without SI practice. To disentangle these two possible accounts we completed a follow-up experiment in which we expanded the sample. Specifically, we recruited an SI-trained group that did have recent practice (students at the end of their coursework for a Master in Conference Interpreting) and an untrained group with no recent practice (students finishing a Triennale in Languages). The four groups considered in the follow-up experiment represent a full crossing of the SI training and recent SI practice factors.

5.5 Follow-up Experiment: Methods

5.5.1 Participants

Two additional groups of participants were recruited for the follow-up experiment. Nineteen students who had recently finished the coursework for a Triennale degree (equivalent to a Bachelor's degree) in Languages from the University of Trieste formed the first group (12 females). The second group consisted of sixteen students who had recently finished the coursework for the Master's program in Conference Interpreting at the University of Trieste (14 females; one participant had attended the Master's program in Conference Interpreting at another university in Italy). All participants were native Italian speakers with no known neurological or psychiatric problems. They all reported normal color vision, which was confirmed with the Ishihara Color Vision Test (Ishihara, 1972). In contrast to the main experiment, participants who spoke a language other than Italian (or Italian dialects) from birth were not included in the experiment. Additionally, the five participants not meeting this criterion from the main experiment were not considered in any analyses. A review of the biographical characteristics of the four groups (including the two from the main experiment) showed no difference between the groups in intelligence (measured with Raven's Advanced Progressive Matrices Series I) or socioeconomic status (measured with mother's years of education; Table 15). Differences in age and years of education were unavoidable given the cross-sectional design of the study. All participants gave written informed consent and were compensated for their time. The study was approved by the ethical committee of the Scuola Internazionale Superiore di Studi Avanzati (SISSA).

Table 15: Biographical and language characteristics of participants in the Follow-up Experiment

	Finishing Triennale (N=16)	Beginning Master's (N=19)	Finishing Master's (N=16)	Writing Thesis (N=19)	*p*-value
Age (in years)	22.4 (1.7)	22.8 (1.8)	24.0 (1.3)	24.5 (1.0)	$p < .001$
Years of education	16.1 (0.7)	16.2 (0.4)	18.2 (0.5)	18.3 (0.7)	$p < .001$
Raven's APM score	10.4 (1.7)	10.7 (1.6)	10.6 (1.8)	11.3 (0.9)	$p = .334$
Mother's years of education	12.1 (3.3)	13.3 (2.6)	14.3 (4.2)	13.3 (3.6)	$p = .423$
Number of functional languages	3.6 (0.6)	3.4 (0.5)	3.6 (0.7)	3.6 (0.8)	$p = .782$
L2 reading	4.3 (0.4)	4.3 (0.5)	4.4 (0.5)	4.7 (0.5)	$p = .014$
L2 writing	3.8 (0.5)	3.9 (0.5)	4.1 (0.3)	4.3 (0.6)	$p = .025$
L2 speaking	3.9 (0.5)	4.1 (0.3)	4.2 (0.4)	4.4 (0.6)	$p = .009$
L2 understanding	4.2 (0.4)	4.2 (0.4)	4.3 (0.4)	4.7 (0.5)	$p = .003$
L3 reading	3.9 (0.3)	3.9 (0.6)	4.2 (0.4)	4.5 (0.6)	$p = .003$
L3 writing	3.4 (0.6)	3.4 (0.6)	3.5 (0.6)	3.7 (0.8)	$p = .496$
L3 speaking	3.2 (0.8)	3.4 (0.7)	3.4 (0.5)	3.6 (0.7)	$p = .367$
L3 understanding	4.1 (0.6)	3.8 (0.5)	4.0 (0.4)	4.3 (0.6)	$p = .088$

Note: Values reported are means with standard deviations in parentheses. For one participant no language data was available and for another participant L2 self-ratings were missing.

5.5.2 Tasks and Procedure

The test session was identical to that used in the main experiment. However, the assignment of L1, L2, and L3 differed (L2 and L3 languages used: Croatian, Dutch, English, French, German, Russian, and Spanish). As in the main experiment, Italian was always assigned to L1. For the Finishing Triennale students, their 1st language of study in their Triennale program was assigned to L2 and their 2nd language of study to L3. For the Finishing Master's students, the assignment was based on the B/C classification of their languages of study. B languages in interpretation are non-native languages which the interpreter uses both actively (interprets from their native language into this language) and passively (interprets from this language into their native language). C languages are non-native languages which are used only passively in interpretation. For students with one B language and one C language, the B language was always assigned to L2 and the C language to L3. For students who had multiple B or C languages, the assignment was identical to the main experiment, that is, according to the participant's confidence in these languages. Given this difference in assignment between the experiments, the languages were later re-coded for each participant based on the average of their self-ratings across the four areas. This led to the switching of L2 and L3 for five participants from the main experiment and four participants from the follow-up experiment. Additionally, participants were asked to report how many hours they had practiced simultaneous interpretation in the two months prior to testing. Three participants in the Finishing Triennale group reported recent practice with simultaneous interpretation and were therefore excluded from all analyses.

The four groups did not differ in terms of their number of functionally fluent languages and self-ratings for L3 writing and speaking (see Table 15 for values).

However, group differences were seen in the self-ratings across the L2 abilities and in L3 reading (*p*s ≤ .025; Table 15; L3 understanding showed a marginal difference). Post hoc Bonferroni-corrected tests revealed that the Thesis group gave higher self-ratings than the Finishing Triennale group on all L2 abilities and L3 reading (*p*s ≤ .028). Additionally, the self-ratings for the Thesis group were higher than those of the Beginning Master's group on L2 reading and understanding and L3 reading (*p*s ≤ .035). No other group differences in self-ratings were significant (*p*s ≥ .070). Finally, the groups did not differ in their frequency of code-switching in all situations (*p*s ≥ .115) except thinking ($p = .002$). Post hoc Bonferroni-corrected tests revealed that the Beginning Master's students code-switched more frequently when thinking than the Finishing Master's and Thesis students ($p = .002$ and $p = .029$, respectively).

5.6 Follow-up Experiment: Results

5.6.1 Verbal Fluency

To analyze differences in verbal fluency, a mixed effects three-way ANOVA with category type (semantic, letter) and language (L1, L2, L3) as within-subjects factors and group (Finishing Triennale, Beginning Master's, Finishing Master's, Thesis) as a between-subjects factor was used. Reported results reflect a Greenhouse-Geisser correction.

As in the main experiment, more responses were given for semantic categories than for letter categories ($F(1,66) = 53.135$, $p < .001$, $\eta_p^2 = .446$; Figure 16) and language affected the number of responses ($F(1.873,123.596) = 221.661$, $p < .001$, $\eta_p^2 = .771$). The greatest number of responses was produced in L1 followed by L2 and then L3 ($ps < .001$ from post hoc t-tests). Unlike the main experiment, there was a main effect of group ($F(3,66) = 2.930$, $p = .040$). Post hoc t-tests revealed that the Finishing Triennale group produced marginally fewer responses than all of the other groups ($ps \leq .026$), with no differences between these groups ($ps \geq .934$). Additionally, category type and language interacted ($F(1.992,131.480) = 6.698$, $p = .002$, $\eta_p^2 = .092$). No other interactions were significant ($ps \geq .201$).

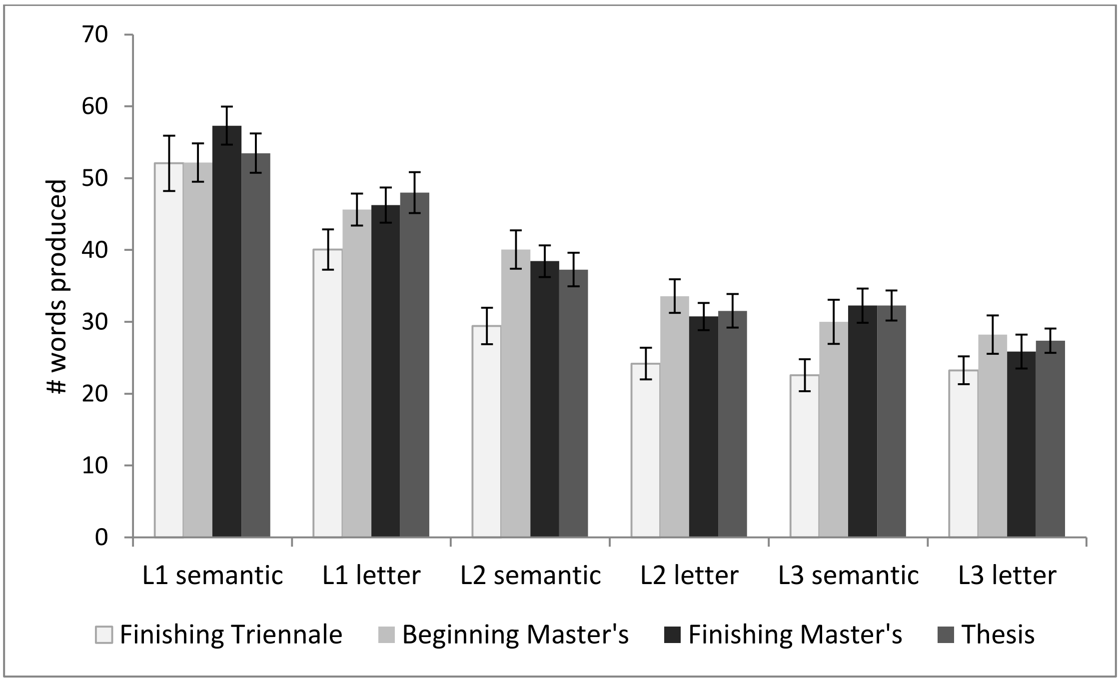

Figure 16: Verbal fluency results for language and category type by group in the Follow-up Experiment. Error bars represent the standard errors of the mean.

5.6.2 Language Switching Task

Data from one participant (Thesis student) was excluded due to a disruption during the task. Non-parametric tests were used to analyze the accuracy data given that these data were not normally distributed. The Kruskal-Wallis test was used to compare between the four groups and the Wilcoxon signed-ranked test was used to compare conditions. The data trimming procedures (12.1% of all trials removed) and analyses performed on the RT data were identical to those in the main experiment.

Analyses on accuracy revealed no overall difference between the groups ($p = .197$). Further, there were no differences between the groups when considering each language and trial type separately (ps $\geq .089$). Across the two groups, accuracy was higher on n-2 non-repetition trials than on n-2 repetition trials (Wilcoxon T = 1227, Z = 2.025, $p = .043$). Additionally, accuracy was higher when responding in L1 than in L2 and L3 (Wilcoxon T = 566, Z = 3.836, $p < .001$ and Wilcoxon T = 654, Z = 3.309, $p = .001$, respectively), with no difference between L2 and L3 ($p = .650$).

The RT analysis showed a main effect of trial type ($F(1,65) = 91.794$, $p < .001$, $\eta_p^2 = .585$) with faster responses to n-2 non-repetition trials than n-2 repetition trials. The main effect of language was also significant ($F(1.831,119.044) = 33.125$, $p < .001$, $\eta_p^2 = .338$). Post hoc t-tests (evaluated at $\alpha = .017$ to correct for multiple comparisons) revealed that responses to L1 were faster than to L2 and L3 ($t(68) = 6.179$, $p < .001$ and $t(68) = 7.879$, $p < .001$, respectively), with no difference between L2 and L3 ($p = .813$). Trial type and language also showed a significant interaction ($F(1.964,127.675) = 8.878$, $p < .001$, $\eta_p^2 = .120$). Through post hoc t-tests (evaluated at $\alpha = .017$ to correct for multiple comparisons), the n-2 repetition cost in L1 was shown to be smaller than in L2 and L3 ($t(68) = 3.045$, $p = .003$ and $t(68) = 3.645$, $p = .001$, respectively), which did not differ ($p = .720$). There was no main effect of group ($p = .533$) or significant interactions between group and either the trial type or language variables ($p = .889$ and $p = .966$, respectively). However, the 3-way interaction of trial type, language, and group was significant ($F(5.839,127.675) = 3.172$, $p = .007$, $\eta_p^2 = .128$; Figure 17). To decompose the 3-way interaction, post hoc one-way ANOVAs (evaluated at $\alpha = .017$ to correct for multiple comparisons) were computed within each language comparing the n-2 repetition cost across the four groups. However, no significant group differences were seen (*p*s ≥ .092).

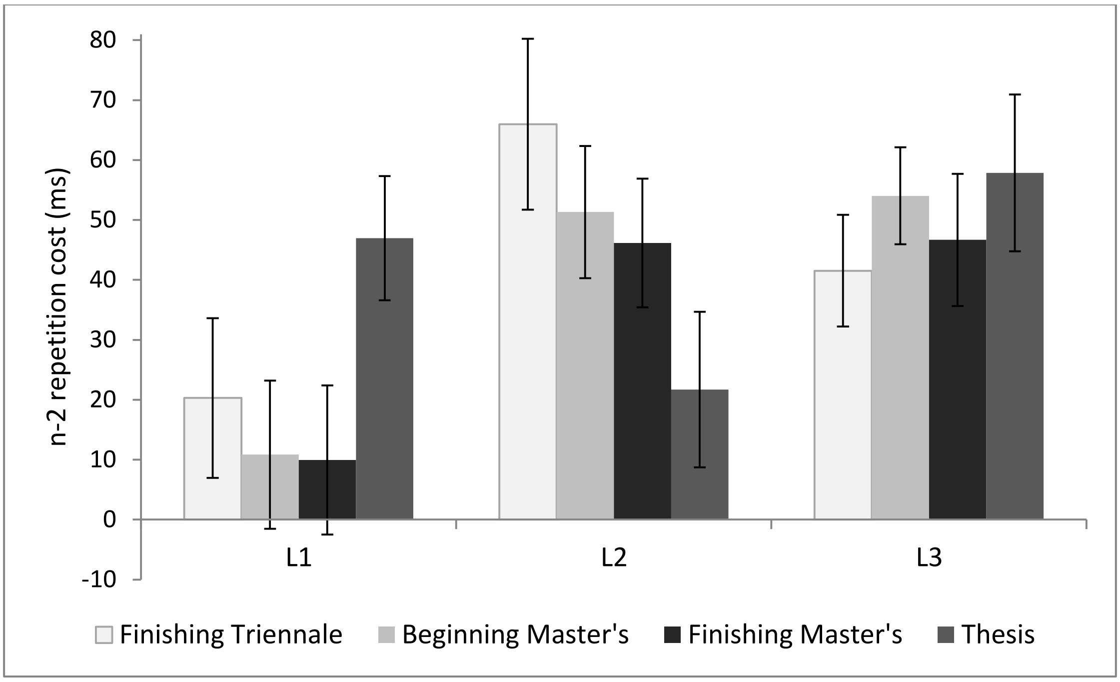

Figure 17: n-2 repetition costs for each language by group on the language switching task in the Follow-up Experiment. Error bars represent the standard errors of the mean.

5.6.2.1 Correlations with language characteristics. As in the main experiment, n-2 repetition costs in L1, L2, and L3 were tested for correlations with the semantic and letter fluency scores in L1, L2, and L3 and self-ratings in L2 and L3. There were no significant correlations between the n-2 repetition costs and any of the verbal fluency measures (*p*s ≥ .232). However, the n-2 repetition cost in the L2 was negatively correlated with self-rated speaking and understanding in L2 ($r = -.237$,

$p = .053$ and $r = -.316$, $p = .009$, respectively), while the n-2 repetition cost in L3 showed a marginal positive correlation with self-rated reading in L3 ($r = .216$, $p = .076$). All other correlations were non-significant ($p\text{s} \geq .107$). The values reported above were not corrected for multiple comparisons as the analyses were of an exploratory nature. Given this fact, these findings should be viewed with caution.

5.6.2.2 Correlations with SI practice. To examine the hypothesis that the different patterns of n-2 repetition costs across the languages seen in the main experiment were due to recent practice of SI, the n-2 repetition costs in L1, L2, and L3 were correlated with the amount of practice values supplied by the Finishing Master's students. Marginal positive correlations between SI practice and the n-2 repetition costs in L2 and L3 were seen ($r = .464$, $p = .070$ and $r = .453$, $p = .078$, respectively).

5.6.3 Non-Language Switching Task

Data from three participants were excluded from all analyses of this task. Two participants (a Beginning Master's student and a Finishing Master's student) were identified as extreme outliers within their groups based on their accuracy rate (more than 3 interquartile ranges below the 1st quartile). An additional participant (Beginning Master's student) was excluded due to difficulties coding her data offline. The exclusion of these participants did not change the results of the comparisons between the two groups on the biographical and language characteristics mentioned above.

Non-parametric tests were used to analyze the accuracy data given that these data were not normally distributed. The Kruskal-Wallis test was used to compare between the four groups and the Wilcoxon signed-ranked test was used to compare conditions. The data trimming procedures (10.4% of all trials removed) and analyses performed on the RT data were identical to those in the main experiment.

Analyses on accuracy revealed no overall difference between the groups (p = .389). Further, there were no differences between the groups when considering each task and trial type separately (ps ≥ .112). Across the two groups, accuracy was higher on n-2 non-repetition trials compared to n-2 repetition trials (Wilcoxon T = 1407, Z = 3.029, p = .002). There were no differences in accuracy between the three tasks (ps ≥ .612). The RT analysis showed a main effect of trial type ($F(1,63) = 419.777$, $p < .001$, $\eta_p^2 = .870$) with faster responses to n-2 non-repetition trials than n-2 repetition trials. No other effects or interactions were significant (ps ≥ .269).

5.7 Follow-up Experiment: Discussion

5.7.1 Summary of Results

This experiment further explored the role of simultaneous interpretation experience on lexical robustness and inhibitory control. Multilingual students with varying levels of SI experience were compared. Lexical robustness differed between the four groups; students with no SI experience had lower verbal fluency scores than the other three groups. Inhibitory control in the language domain also differed across the four groups, though it was the group with greatest amount of training that displayed an altered pattern in this case. Inhibitory control in the non-language domain, however, was unaffected by SI experience.

5.7.2 Lexical Robustness

While lexical robustness is likely related to successful simultaneous interpretation, this characteristic does not appear to be influenced by either training in SI or recent SI practice. Instead, a division was seen between students who had been accepted to and enrolled in a Master in Conference Interpreting program and those who had not. Thus, it appears that high lexical robustness is a prerequisite for simultaneous interpretation, or at least acceptance into a training program. This, however, does not exclude verbal fluency from being a trained ability; in fact, a hint of training does exist in the data. The Finishing Triennale students will become the next pool of Master's program applicants, and presumably those admitted will have higher verbal fluencies, on par with the three other groups, than those not admitted. If verbal fluency was not trained between our study and the admissions exam this would be seen as greater variability of individual scores in the Finishing Triennale group than in the other three groups. This difference in variability is not seen in the present data. Further, it is highly likely that applicants to SI training programs work to improve their languages in the months before the admissions exam. Indeed, common advice to aspiring interpreters is to improve their native language. Though interestingly, the present data show larger differences in L2 verbal fluency, not L1.

While it appears that verbal fluency could be a useful tool in selecting between Master's program applicants, more research should be completed to flesh out the role it plays. In particular, studies comparing verbal fluency with outcomes on currently used admissions exams and later test scores and progress would be highly informative. It should also be noted that the verbal fluency task is a measure of both vocabulary size and access to items. Arguably, the latter of these two components is more directly connected to successful simultaneous interpretation. Future research may wish to examine the independent contributions of these two components to SI performance.

5.7.3 Inhibitory Control

The follow-up experiment was intended to differentiate between two possible explanations of the language switching data from the main experiment. The first explanation posited that training in SI leads to an increase in inhibition applied to the L1. Support for this account is not present in the current data as the trained Finishing Master's students evidenced the same pattern of results as the untrained Finishing Triennale and Beginning Master's students. The second explanation put forward suggested that recent SI practice, not training, leads to a decrease in inhibition applied to the L1. Support for this account is also lacking. The Finishing Triennale students who were untrained and had no practice with SI patterned with the Beginning Master's and Finishing Master's groups, both of which had recent SI practice.

Though the results of the follow-up experiment do not appear to support either previously proposed account, the data do reveal two patterns of n-2 repetition costs across the languages. The Thesis group showed one pattern in which similar levels of inhibition were applied to L1 and L3, with potentially less inhibition applied to L2. While the other three groups showed a pattern of smaller n-2 repetition costs in the L1 than in the L2 and L3, which did not differ. As the Thesis group is not unique among the four groups in either SI training or practice, the cause of this difference likely lies in other language or experience characteristics.

Reviewing the self-rated proficiency levels of the four groups, the Thesis group emerges as the most highly proficient in the L2. Thus, the different pattern seen among the Thesis students may be related to this increased proficiency. An effect of

L2 proficiency was seen in the L2 n-2 repetition cost, where higher oral proficiency was associated with less inhibition (though this result should be viewed with caution). However, no such association was present in the L1 where group differences in n-2 repetition cost were also apparent.

In addition to L2 proficiency, the Thesis students may differ from the other groups in their prevalent language context. The three non-Thesis groups were all actively engaged in courses and school-life at the time of testing. The Thesis students, on the other hand, were engaged in the solitary task of thesis writing. Notably, the frequency of code-switching across the groups was highest for the "at school" situation. Thus, if the Thesis students are spending considerably less time in the "at school" situation, their balance between code-switching exchanges and single- or dual-language exchanges may have shifted toward the latter. As languages in the single- and dual-language contexts are thought to be in competition, this shift in context may have caused greater inhibition among the Thesis students. This explanation can account for the greater L1 inhibition seen the Thesis students, though similar changes in L2 inhibition were not present. Thus, a combination of L2 proficiency and language context may have contributed to the differential patterns seen across the groups.

This interaction of factors may also explain the variety of patterns seen in the previous studies of the n-2 repetition cost in language switching. Comparing the present patterns with those studies, a similarity is evident between the three non-Thesis groups and the visually noted smaller cost in L1 in Philipp and Koch (2009). Additionally, the smaller cost in L2 seen in the Thesis students is mirrored by the participants in Philipp and colleagues (2007). The previous studies, however, do not provide sufficient information about language context and proficiency to judge the comparability of these groups.

Future examinations of the n-2 repetition cost in language switching would increase our knowledge of the specific role that inhibitory control plays in language management. Unlike switching costs, which may reflect both inhibition and activation (Koch & Philipp, 2005), the n-2 repetition cost is a pure measure of inhibition (Kiesel et al., 2010; Koch et al., 2010). Thus, this measure provides a tool to examine inhibitory control in varying types of multilinguals.

Finally, inhibitory control in the non-language domain appears to be unaffected by these language use factors. However, given that no overall difference in inhibitory control was seen on the language task, but rather the differences were specific to the language classification, it is not surprising that differences did not exist on the non-language task.

5.7.4 Limitations and Future Directions

The present study set out to understand the use of inhibitory control in simultaneous interpretation and instead provided initial information about inhibitory control use in multilinguals more generally. As the study was not designed for this latter question, some pertinent information may be missing, such as more specific estimates of code-switching usage and common language contexts. Future studies may address this by creating a more specialized language questionnaire.

Regarding the initial aim of the study, the use of students proved to complicate the results, as the primary language context of students appears to be driven not by experience with simultaneous interpretation, but rather the educational environment.

Future studies may better target the role of inhibition in simultaneous interpretation by examining professional interpreters using the n-2 repetition cost measure. These studies should also employ broad language questionnaires to consider all possible influences.

Finally, the current paradigm made use of a productive language switching paradigm, however, the simultaneity of languages in SI may apply specifically to receptive language (see Grosjean, 1997 for a model). Thus, an investigation using a receptive language switching paradigm may furnish different results.

5.8 Concluding Remarks

The present study demonstrated that the level of inhibitory control used in language management is dependent on various language and language use characteristics of the individual bilingual. Further, it promoted the use of the n-2 repetition cost measure in studying inhibitory control among multilinguals. Finally, enhanced verbal fluency appears to be related to acceptance to a simultaneous interpretation training program.

CHAPTER 6

THE FINAL INTERPRETATION

The present project set out to examine the fingerprint that simultaneous interpretation leaves on the mind and brain. By investigating the neurocognitive profile of professional interpreters and students of interpretation, we hope to provide further insight into the complex processes that make up the simultaneous interpretation skill. The data presented here support the narrative that interpreters are not expert bilinguals, but rather possess enhancements that are specific to the needs of simultaneous interpretation. This finding, which is in line with a recent study (Morales et al., 2015), suggests that simultaneous interpretation is not merely an extreme form of bilingualism, but a skill that builds upon bilingualism. The data further suggest that the enhancements seen in interpreters are due to a combination of innate abilities and trained abilities.

The abilities which appeared to be the precursors to successful acquisition of the simultaneous interpretation skill clustered around language-related functions. These functions included verbal intelligence, verbal fluency, strength of the language white matter connections, and efficiency of the spoken word learning network. Thus, it appears that strong verbal abilities may be a prerequisite for acceptance into an SI training program. Indeed, anecdotal evidence suggests that this is true; aspiring interpreters are often given the advice to strengthen their native language, which may particularly affect verbal fluency. This advice also highlights the fact that “innate” is likely a misnomer for these abilities. A more accurate label may be “acquired before training.” This cluster of abilities further underwrites the importance of high verbal aptitude in simultaneous interpretation.

In contrast to the “innate” abilities, SI training evidenced growth in several cognitive control abilities, in particular, verbal and spatial short-term memory, verbal working memory, and brain regions involved in the simultaneity of input and output. These changes may be taken as evidence of the processes that are particularly recruited during simultaneous interpretation and therefore received targeted training. The use of verbal memory is line with the need to store source language input and pre-output target language utterances. The enhancement in spatial short-term memory further suggests that simultaneous interpretation may additionally rely on some types of non-verbal memory. Finally, the growth in brain regions which support the simultaneity of speaking and listening pinpoints this aspect of simultaneous interpretation as particularly difficult and unique to SI.

Notably, training or practice with simultaneous interpretation did not appear to influence the use of inhibitory control as a language management mechanism. Evidence for this lack of change comes primarily from the examination of n-2 repetition costs; however, it is corroborated by the absence of changes in gray matter volume in regions supporting language control. These data suggest that simultaneous interpretation makes use of the same options for language control as those available to non-interpreter multilinguals.

The findings described in the presented studies may be of particular use to simultaneous interpretation training programs. These data may assist in creating admissions procedures and training materials which target the skills required for SI. It is important to note, however, that the division of “innate” and trained abilities detailed above may be the result of already present admissions and training procedures. Specifically, admissions tests may assess verbal abilities, but not general cognitive abilities. This has two implications. First, admissions procedures could potentially be improved through the assessment of additional abilities. Second, the labeling of verbal abilities as a prerequisite may be premature, as the differences seen may be a reflection of the current selection procedures. Future studies should examine a broader set of applicants to adjudicate these issues.

The studies presented in this dissertation represent an initial examination of the effect that the complex task of simultaneous interpretation has on the mind and brain. The findings have direct significance for simultaneous interpretation trainers and researchers, as well as for understanding language processing more generally. Additionally, the results speak to the extraordinary plasticity of the brain during skill learning even in adulthood. Future research on simultaneous interpretation will further our understanding of the intricacies and immense capacity of the human brain.

ACKNOWLEDGEMENTS

I could not have completed the work presented in this dissertation without the help and support of numerous others. They all have my deep gratitude for assisting me along my path to this point. First and foremost among them is my supervisor, Antonio Vallesi. Though he was at first unsure that our research interests would intersect, I am grateful that he took the time to develop a research program that suits us both so well. I am additionally indebted to my labmate, Olga Puccioni, for helping with task designs, collecting data, correcting my Italian, and for being a friend through these years. I am grateful also to Loredana Torrisi who was the impetus for the fourth study and collected half of the data.

Further I would like to thank Alessio Isaja for providing constant technical support and Laura Riontino for assisting in data collection and participant recruitment. Additionally, I am grateful to Marta Maieron and the technicians in Udine, Antonio Margiotta, Mauro Molinaro, Filippo Cotrer, and Cristian Piorico, for all their help with the acquisition of the neuroimaging data. I would also like to thank Regine Kolinsky and the UNESCOG lab group for giving me a home in Brussels; and Ian Anderson and Lucinda Afreixo for helping me contact the professional interpreters that made my time there a success. Further, Maurizio Viezzi, Daniela Negru, and Federica Scarpa assisted immeasurably in the recruitment of the interpretation and translation students. Additionally, Randall Engle welcomed me to his lab group and provided the working memory tasks.

A special thanks goes to the other students in my year, Ana Laura, Jenny, Vahid, and Vanni, without the support of the little "family" we created our first year I don't know that I would have continued. Numerous other friends at SISSA have made provided support both related to and unrelated to science.

Finally, I could not have done any of these studies without my amazing participants. I am eternally grateful to them for participating, being genuinely interested in the study, travelling to Udine, and improving my Italian. They truly are the superstars of this project.

REFERENCES

Abutalebi, J., Canini, M., Della Rosa, P. A., Sheung, L. P., Green, D. W., & Weekes, B. S. (2014). Bilingualism protects anterior temporal lobe integrity in aging. *Neurobiology of Aging*, *35*(9), 2126–2133. doi:10.1016/j.neurobiolaging.2014.03.010

Aherns, B., Kalderon, E., Krick, C. M., & Reith, W. (2010). fMRI for exploring simultaneous interpreting. In D. Gile, G. Hansen, & N. K. Pokorn (Eds.), *Why Translation Studies Matter* (pp. 237–248). Amsterdam: John Benjamins.

Anderson, L. (1994). Simultaneous interpretation: Contextual and translational aspects. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the Gap: Empirical research in simultaneous interpretation* (pp. 101–120). Amsterdam: Benjamins.

Arbuthnott, K. (2008). The effect of task location and task type on backward inhibition. *Memory & Cognition*, *36*(3), 534–543. doi:10.3758/MC.36.3.534

Arbuthnott, K., & Frank, J. (2000). Executive control in set switching : Residual switch cost and task-set inhibition. *Canadian Journal of Experimental Psychology*, *54*(1), 33-41.

Babcock, L., & Vallesi, A. (2015). The interaction of process and domain in prefrontal cortex during inductive reasoning. *Neuropsychologia*, *67*, 91–99. doi:10.1016/j.neuropsychologia.2014.12.010

Bajo, M. T., Padilla, F., & Padilla, P. (2000). Comprehension processes in simultaneous interpreting. In A. Chesterman, N. Gallardo San Salvador, & Y. Gambier (Eds.), *Translation in Context* (pp. 127–142). Amsterdam: John Benjamins.

Berkman, E. T., Kahn, L. E., & Merchant, J. S. (2014). Training-induced changes in inhibitory control network activity. *Journal of Neuroscience*, *34*(1), 149–157. doi:10.1523/JNEUROSCI.3564-13.2014

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology*, *60*(1), 68–79. doi:10.1037/cjep2006008

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, *10*(3), 89–129. doi:10.1177/1529100610387084

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–249. doi:10.1016/j.tics.2012.03.001

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 859–73. doi:10.1037/0278-7393.34.4.859

Bialystok, E., & Depape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 565–74. doi:10.1037/a0012735

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*(1), 103–119. doi:10.1177/13670069050090010701

Branzi, F. M. (2014). *The mechanisms and the scope of bilingual language production*. (Unpublished doctoral dissertation). University Pompeu Fabra.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, *39*(4), 713–726.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*, *25*, 958–968.

Burns, J. T. (1965). *The effect of errors on reaction time in a serial reaction task*. (Unpublished doctoral dissertation). University of Michigan.

Carreiras, M., Seghier, M. L., Baquero, S., Estévez, A., Lozano, A., Devlin, J. T., & Price, C. J. (2009). An anatomical signature for literacy. *Nature*, *461*, 983–986. doi:10.1038/nature08461

Catani, M., Allin, M. P. G., Husain, M., Pugliese, L., Mesulam, M. M., Murray, R. M., & Jones, D. K. (2007). Symmetries in human brain language pathways correlate with verbal recall. *PNAS*, *104*(23), 17163–17168. doi:10.1073/pnas.0702116104

Catani, M., Dell'Acqua, F., Vergani, F., Malik, F., Hodge, H., Roy, P., … Thiebaut de Schotten, M. (2012). Short frontal lobe connections of the human brain. *Cortex*, *48*(2), 273–291. doi:10.1016/j.cortex.2011.12.001

Catani, M., Howard, R. J., Pajevic, S., & Jones, D. K. (2002). Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *NeuroImage*, *17*(1), 77–94. doi:10.1006/nimg.2002.1136

Catani, M., Jones, D. K., & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, *57*(1), 8–16. doi:10.1002/ana.20319

Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: History and current state. *Cortex*, *44*(8), 953–961. doi:10.1016/j.cortex.2008.04.002

Catani, M., Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., … Rogalski, E. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain*, *136*, 2619–2628. doi:10.1093/brain/awt163

Catani, M., & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, *44*(8), 1105–1132. doi:10.1016/j.cortex.2008.05.004

Christoffels, I. K., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*(3), 324–345. doi:10.1016/j.jml.2005.12.004

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*(2), 135–149. doi:10.1016/j.cognition.2009.08.001

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*(1), 59–86. doi:10.1016/j.cognition.2006.12.013

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1057–1074. doi:10.1037/0278-7393.32.5.1057

Darò, V. (1989). The role of memory and attention in simultaneous interpretation: A neurolinguistic approach. *The Interpreters' Newsletter*, *2*, 50–56.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*, 3773–3800. doi:10.1098/rstb.2009.0111

De Groot, A. M. B., & Christoffels, I. K. (2006). Language control in bilinguals: Monolingual tasks and simultaneous interpreting. *Bilingualism: Language and Cognition*, *9*(2), 189–201. doi:10.1017/S1366728906002537

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*, 496–518.

Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist Connectist Approaches to Human Cognition* (pp. 189–226). Mahwah, NJ: Lawrence Erlbaum.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Changes in grey matter induced by training. *Nature*, *427*, 311–312. doi:10.1038/427311a

Draganski, B., & May, A. (2008). Training-induced structural changes in the adult human brain. *Behavioral Brain Research*, *192*(1), 137–142. doi:10.1016/j.bbr.2008.02.015

Elmer, S., Hänggi, J., & Jäncke, L. (2014). Processing demands upon cognitive, linguistic, and articulatory functions promote grey matter plasticity in the adult multilingual brain: Insights from simultaneous interpreters. *Cortex*, *54*, 179–189. doi:10.1016/j.cortex.2014.02.014

Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2011). Differential language expertise related to white matter architecture in regions subserving sensory-motor coupling, articulation, and interhemispheric transfer. *Human Brain Mapping*, *32*, 2064–2074. doi:10.1002/hbm.21169

Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science*, *19*(12), 1201–1206. doi:10.1111/j.1467-9280.2008.02224.x

Fabbro, F., & Gran, L. (1994). Neurological and neuropsychological aspects of polyglossia and simultaneous interpretation. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the Gap: Empirical research in simultaneous interpretation* (pp. 273–317). Amsterdam: Benjamins.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135.

Fulbright, R. K., Manson, S. C., Skudlarski, P., Lacadie, C. M., & Gore, J. C. (2003). Quantity determination and the distance effect with letters, numbers, and shapes: A functional MR imaging study of number processing. *American Journal of Neuroradiology*, *24*, 193–200.

Gade, M., & Koch, I. (2012). Inhibitory processes for critical situations – the role of n−2 task repetition costs in human multitasking situations. *Frontiers in Physiology*, *3*, 159. doi:10.3389/fphys.2012.00159

Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., … Ávila, C. (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *NeuroImage*, *53*(4), 1272–1278. doi:10.1016/j.neuroimage.2010.05.078

Gaser, C., & Schlaug, G. (2003). Gray matter differences between musicians and nonmusicians. *Annals of the New York Academy of Sciences*, *999*, 514–517. doi:10.1196/annals.1284.062

Ge, Y., Grossman, R. I., Babb, J. S., Rabin, M. L., Mannon, L. J., & Kolson, D. L. (2002). Age-related total gray matter and white matter changes in normal adult brain. Part I: volumetric MR imaging analysis. *American Journal of Neuroradiology*, *23*, 1327–1333.

Geschwind, N. (1967). Wernicke's contribution to the study of aphasia. *Cortex, 3*, 449–463.

Gottfried, A. W., Gottfried, A. E., Bathurst, K., Guerin, D. W., & Parramore, M. M. (2003). Socioeconomic status in children's development and family environment: Infancy through adolescence. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic Status, Parenting, and Child Development* (pp. 189–207). Mahwah, NJ: Lawrence Erlbaum.

Gran, L., & Fabbro, F. (1988). The role of neuroscience in the teaching of interpretation. *The Interpreters' Newsletter*, *1*, 23–41.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67–81.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530. doi:10.1080/20445911.2013.796377

Grieve, S. M., Williams, L. M., Paul, R. H., Clark, C. R., & Gordon, E. (2007). Cognitive aging, executive function, and fractional anisotropy: A diffusion tensor MR imaging study. *American Journal of Neuroradiology*, *28*, 226–235. doi:28/2/226 [pii]

Grosjean, F. (1997). The bilingual individual. *Interpreting*, *2*, 163–187.

Guo, T., Liu, F., Chen, B., & Li, S. (2013). Inhibition of non-target languages in multilingual word production: Evidence from uighur-chinese-english trilinguals. *Acta Psychologica*, *143*(3), 277–283. doi:10.1016/j.actpsy.2013.04.002

Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition*, *13*(03), 315–325. doi:10.1017/S1366728909990010

Hernández, M., Martin, C. D., Barceló, F., & Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language*, *69*(3), 257–276. doi:10.1016/j.jml.2013.06.004

Hervais-Adelman, A., Moser-Mercer, B., Michel, C. M., & Golestani, N. (2014). fMRI of simultaneous interpretation reveals the neural basis of extreme language control. *Cerebral Cortex*. Advance online publication. doi:10.1093/cercor/bhu158

Houghton, G., Pritchard, R., & Grange, J. A. (2009). The role of cue-target translation in backward inhibition of attentional set. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 466–476. doi:10.1037/a0014648

Ibáñez, a J., Macizo, P., & Bajo, M. T. (2010). Language access and language selection in professional translators. *Acta Psychologica*, *135*(2), 257–66. doi:10.1016/j.actpsy.2010.07.009

Imfeld, A., Oechslin, M. S., Meyer, M., Loenneker, T., & Jancke, L. (2009). White matter plasticity in the corticospinal tract of musicians: A diffusion tensor imaging study. *NeuroImage*, *46*(3), 600–607. doi:10.1016/j.neuroimage.2009.02.025

Ishihara, S. (1972). *Tests for colour-blindness*. Tokyo, Japan: Kanehara Shuppan Co., Ltd.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. doi:10.1037/0096-3445.133.2.189

Kavé, G., Eyal, N., Shorek, A., & Cohen-Mansfield, J. (2008). Multilingualism and cognitive state in the oldest old. *Psychology and Aging*, *23*(1), 70–78. doi:10.1037/0882-7974.23.1.70

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching--a review. *Psychological Bulletin*, *136*(5), 849–874. doi:10.1037/a0019842

Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching: a review. *Psychonomic Bulletin & Review*, *17*(1), 1–14. doi:10.3758/PBR.17.1.1

Koch, I., & Philipp, A. M. (2005). Effects of response selection on the task repetition benefit in task switching. *Memory & Cognition*, *33*(4), 624–634.

Koch, I., Prinz, W., & Allport, A. (2005). Involuntary retrieval in alphabet-arithmetic tasks: Task-mixing and task-switching costs. *Psychological Research*, *69*(4), 252–261. doi:10.1007/s00426-004-0180-y

Köpke, B., & Nespoulous, J.-L. (2006). Working memory performance in expert and novice interpreters. *Interpreting*, *8*(1), 1–23.

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging*, *15*(1), 126–147. doi:10.1037//OS82-797-U5.1.126

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, *128*(3), 416–430. doi:10.1016/j.actpsy.2008.02.001

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*(2), 119–135. doi:10.1017/S1366728906002483

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, mind, and brain. *Annual Review of Linguistics*, *1*, 377–394. doi:10.1146/annurev-linguist-030514-124937

Lewand, R. E. (2000). *Cryptological Mathematics* (p. 199). Washington, DC: The Mathematical Association of America.

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

Lu, L. H., Crosson, B., Nadeau, S. E., Heilman, K. M., Gonzalez-rothi, L. J., Raymer, A., … Roper, S. N. (2002). Category-specific naming deficits for objects and actions: semantic attribute and grammatical role hypotheses. *Neuropsychologia*, *40*, 1608–1621.

Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, *14*(4), 596–603. doi:10.1017/S1366728911000186

Macnamara, B. N., & Conway, A. R. A. (2013). Novel evidence in support of the bilingual advantage: Influences of task demands and experience on cognitive control and working memory. *Psychonomic Bulletin & Review*, *21*(2), 520–525. doi:10.3758/s13423-013-0524-y

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *PNAS*, *97*(8), 4398–4403.

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, *86*(1), 70–82. doi:10.1016/S0093-934X(02)00535-7

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, *63*(1), 240–244. doi:10.1016/j.neuroimage.2012.06.043

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, *129*(1), 4–26. doi:10.10371/0096-3445.129.1.4

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1423–1442.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–91. doi:10.1037/a0028228

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*, 25–40.

Michael, E. B., & Gollan, T. H. (2005). Being and becoming bilingual: Individual differences and consequences for langauge production. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholingusitic Approaches* (pp. 389–407). Oxford: Oxford University Press.

Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y. T., … Nestor, P. J. (2010). What the left and right anterior fusiform gyri tell us about semantic memory. *Brain*, *133*, 3256–3268. doi:10.1093/brain/awq272

Morales, J., Padilla, F., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Simultaneous interpretation selectively influences working memory and attentional networks. *Acta Psychologica*, *155*, 82–91. doi:10.1016/j.actpsy.2014.12.004

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220–246.

Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science*, *10*(4), 464–80. doi:10.1111/j.1467-7687.2007.00600.x

Orsini, A., & Laicardi, C. (1997). *WAIS-R. Contributo alla taratura italiana* (p. 120). Giunti Organizzazioni Speciali.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258. doi:10.1016/j.cogpsych.2012.12.002

Padilla, F., Bajo, M. T., & Macizo, P. (2005). Articulatory suppression in language interpretation: Working memory capacity, dual tasking and word knowledge. *Bilingualism: Language and Cognition*, *8*(3), 207–219. doi:10.1017/S1366728905002269

Padilla, P., Bajo, M. T., Cañas, J. J., & Padilla, F. (1995). Cognitive processes of memory in simultaneous interpretation. In J. Tommola (Ed.), *Topics in Interpreting Research*. Turku, Finland: University of Turku Centre for Translation and Interpreting.

Paradis, M. (1994). Toward a neurolinguistic theory of simultaneous translation: The framework. *International Journal of Psycholinguistics*, *10*, 319–335.

Perani, D. (2005). The neural basis of language talent in bilinguals. *Trends in Cognitive Sciences*, *9*(5), 211–213. doi:10.1016/j.tics.2005.03.001

Peschke, C., Ziegler, W., Kappes, J., & Baumgaertner, a. (2009). Auditory-motor integration during fast repetition: The neuronal correlates of shadowing. *NeuroImage*, *47*(1), 392–402. doi:10.1016/j.neuroimage.2009.03.061

Philipp, A. M., Gade, M., & Koch, I. (2007). Inhibitory processes in language switching: Evidence from switching language-defined response sets. *European Journal of Cognitive Psychology*, *19*(3), 395–416.

Philipp, A. M., & Koch, I. (2006). Task inhibition and task repetition in task switching. *European Journal of Cognitive Psychology*, *18*(4), 624–639. doi:10.1080/09541440500423269

Philipp, A. M., & Koch, I. (2009). Inhibition in language switching: What is inhibited when switching between languages in naming tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1187–1195. doi:10.1037/a0016376

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*(4), 535–51. doi:10.1016/j.jecp.2012.06.013

Prior, A. (2012). Too much of a good thing: Stronger bilingual inhibition leads to larger lag-2 task repetition costs. *Cognition*, *125*(1), 1–12. doi:10.1016/j.cognition.2012.06.019

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society*, *17*(04), 682–691. doi:10.1017/S1355617711000580

Prior, A., & Macwhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*(2), 253–262. doi:10.1017/S1366728909990526

Puccioni, O., & Vallesi, A. (2012a). High cognitive reserve is associated with a reduced age-related deficit in spatial conflict resolution. *Frontiers in Human Neuroscience*, *6*, 327. doi:10.3389/fnhum.2012.00327

Puccioni, O., & Vallesi, A. (2012b). Sequential congruency effects: disentangling priming and conflict adaptation. *Psychological Research*, *76*(5), 591–600. doi:10.1007/s00426-011-0360-5

Ramón y Cajal, S. (1999). *Advice for a Young Investigator* (p. 150). MIT Press.

Raven, J., Raven, J. C., & Court, J. H. (1998). Section 4: The Advanced Progressive Matrices. In *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessement.

Rinne, J. O., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., … Sunnari, M. (2000). The translating brain: Cerebral activation patterns during simultaneous interpreting. *Neuroscience Letters*, *294*(2), 85–88.

Rodriguez-Fornells, A., van der Lugt, A., Rotte, M., Britti, B., Heinze, H.-J., & Münte, T. F. (2005). Second language interferes with word production in fluent bilinguals: Brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience*, *17*(3), 422–433.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231.

Rubin, O., & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1477–1491. doi:10.1037/0278-7393.31.6.1477

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, *8*(4), 535–543.

Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science*, *14*(6), 317–320. doi:10.1111/j.0963-7214.2005.00389.x

Schuch, S., & Koch, I. (2003). The role of response selection for inhibition of task sets in task shifting. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 92–105. doi:10.1037/0096-1523.29.1.92

Schwieter, J. W., & Sunderman, G. (2011). Inhibitory control processes and lexical access in trilingual speech production. *Linguistic Approaches to Bilingualism*, *1*(4), 391–412. doi:10.1075/lab.1.4.02sch

Seeber, K. G., & Kerzel, D. (2011). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism*, *16*(2), 228–242. doi:10.1177/1367006911402982

Signorelli, T. M., Haarmann, H. J., & Obler, L. K. (2011). Working memory in simultaneous interpreters: Effects of task and age. *International Journal of Bilingualism*, *16*(2), 198–212. doi:10.1177/1367006911403200

Singh, S., & Galli, S. (1999). *Codici e Segreti*. Milano: Rizzoli.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281–284. doi:10.1111/1467-9280.00151

Stavrakaki, S., Megari, K., Kosmidis, M. H., Apostolidou, M., & Takou, E. (2012). Working memory and verbal fluency in simultaneous interpreters. *Journal of*

Clinical and Experimental Neuropsychology, *34*(6), 624–33. doi:10.1080/13803395.2012.667068

Stein, M., Federspiel, A., Koenig, T., Wirth, M., Strik, W., Wiest, R., … Dierks, T. (2012). Structural plasticity in the language system related to increased second language proficiency. *Cortex*, *48*(4), 458–465. doi:10.1016/j.cortex.2010.10.007

Steinhauser, M., & Hübner, R. (2005). Mixing costs in task shifting reflect sequential processing stages in a multicomponent task. *Memory & Cognition*, *33*(8), 1484–1494.

Stevens, C., Lauinger, B., & Neville, H. (2009). Differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds: an event-related brain potential study. *Developmental Science*, *12*(4), 634–46. doi:10.1111/j.1467-7687.2009.00807.x

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Tao, L., Marzecová, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: the role of age of acquisition. *Frontiers in Psychology*, *2*, 123. doi:10.3389/fpsyg.2011.00123

Thiebaut de Schotten, M., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D. G. M., & Catani, M. (2011). A lateralized brain network for visuospatial attention. *Nature Neuroscience*, *14*(10), 1245–1246. doi:10.1038/nn.2905

Tzou, Y.-Z., Eslami, Z. R., Chen, H.-C., & Vaid, J. (2011). Effect of language proficiency and degree of formal training in simultaneous interpreting on working memory and interpreting performance: Evidence from Mandarin-English speakers. *International Journal of Bilingualism*, *16*(2), 213–227. doi:10.1177/1367006911403197

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*(4), 780–789.

Van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex*, *18*(11), 2706–2716. doi:10.1093/cercor/bhn030

Walhovd, K. B., Fjell, A. M., Reinvang, I., Lundervold, A., Dale, A. M., Eilertsen, D. E., … Fischl, B. (2005). Effects of age on volumes of cortex, white matter and subcortical structures. *Neurobiology of Aging*, *26*, 1261–1270. doi:10.1016/j.neurobiolaging.2005.05.020

Wechsler, R. (1981). *WAIS-R Manual*. New York: Psychological Corporation.

Yudes, C., Macizo, P., & Bajo, T. (2011). The influence of expertise in simultaneous interpreting on non-verbal executive processes. *Frontiers in Psychology*, *2*, 309. doi:10.3389/fpsyg.2011.00309

Yudes, C., Macizo, P., & Bajo, T. (2012). Coordinating comprehension and production in simultaneous interpreters: Evidence from the Articulatory Suppression Effect. *Bilingualism: Language and Cognition*, *15*(2), 329–339. doi:10.1017/S1366728911000150

Appendix A: Functionally Fluent Questionnaire

English version

Please list all the foreign languages you have studied or know. For each language please answer the following questions with that language in mind.

1. Do you know the A1, A2, B1 etc. system of classifying language levels? If yes, what is your level in this language?
2. Could you discuss a topic in which you are not an expert, such as politics, in this language?
3. Can you understand news programs in this language?
4. Could you read a novel or short story for pleasure in this language?
5. Could you tell a story about events in the past, present, and future to a group of people in this language?
6. Could you write a letter in this language to a friend about an important event in your life and how it affected you?
7. Would you be able to understand an announcement about a cancelled train in this language and follow the directions given about where to refund or change your ticket?
8. Could you write an essay in this language on a work of literature?
9. Could you understand a textbook passage on your field of study in this language?

Italian version

Per favore elenca le lingue straniere che hai studiato o che conosci. Per ogni lingua, per favore rispondi alle seguenti domande riferendoti a quella lingua.

1. Conosci il sistema di classificazione A1, A2, B2, ecc. che indica il livello di conoscenza di una lingua? Se si, quale e’ il tuo livello in questa lingua?
2. Potresti parlare di un tema del quale non sei esperto, come la politica, in questa lingua?
3. Potresti comprendere un telegiornale in questa lingua?
4. Potresti leggere un romanzo o un breve racconto per passatempo in questa lingua?
5. Potresti raccontare una storia su un evento del passato, uno del presente e uno del futuro a un gruppo di persone in questa lingua?
6. Potresti scrivere una lettera in questa lingua ad un amico riguardo un evento importante della tua vita e di come ti ha influenzato?
7. Saresti capace di comprendere un annuncio riguardo un treno cancellato in questa lingua e seguire le istruzioni date su dove farti rimborsare o cambiare il biglietto?
8. Potresti scrivere un saggio in questa lingua su un brano di letteratura?
9. Potresti comprendere un passaggio di un libro di testo nel tuo campo di interesse in questa lingua?

Appendix B: Verbal Fluency Categories in Each Language

Set	English	Italian	Croatian	Dutch	French	German	Russian	Spanish
Semantic set 1	body parts	parti del corpo	dijelovi tijela	lichaamsdelen	parties du corps	Körperteile	части тела	partes del cuerpo
	occupations	professioni	zanimanja	beroepen	métiers	Berufe	профессии	profesiones
	musical instruments	strumenti musicali	glazbeni instrumenti	muziekinstrumenten	instruments de musique	Musikinstrumente	музыкальные инструменты	instrumentos musicales
Semantic set 2	animals	animali	životinje	dieren	animaux	Tiere	животные	animales
	vegetables	legumi	biljke	groente	légumes	Gemüse	овощи	legumbres
	clothing items	abbigliamento	odjeća	kledingstukken	vêtements	Kleidungsstücke	виды одежды	ropa
Semantic set 3	methods of transportation	mezzi di trasporto	prijevozna sredstva	vervoermiddelen	transports	Verkehrsmittel	виды транспорта	medios de transporte
	sports	sport	sportovi	sporttaken	sport	Sportarten	виды спорта	deportes
	fruits	frutta	voće	fruit	fruits	Obst	фрукты	fruta
Letter set 1	A M F	A M F	A M F	A M F	A M F	A M F	А М Ф	A M F
Letter set 2	E T G	E T G	E T G	E T G	E T G	E T G	Е Т Г	E T G
Letter set 3	O D L	O D L	O D L	O D L	O D L	O D L	О Д Л	O D L